Round 1

Reviewer A:

Review of "A quasi-real-time system for automatic local event monitoring in Germany" by Ramos et al.

Summary:
Study topic/area – Germany, automatic real-time event localization/refinement
Main goal(s) – design an automatic system for refining earthquake hypocenters, particularly depths, and assessing their quality/reliability
Methods/analyses – TieBeNN (PhaseNet picker, PyOcto associator, NonLinLoc), Location Quality Score, catalog comparisons, synthetic events
Results – auto depths generally showed good accuracy with higher uncertainty typically corresponding to lower quality values, LQS wasn't as meaningful for depths <~30 km
Interpretations/conclusions – TieBeNN has provided benefits to routine monitoring operations, additional adjustments and development could further improve its performance

This manuscript describes the development of a near real-time/post-processing pipeline to refine events detected by the EdB's existing real-time system, particularly by adding depths, as well as assigning a quality value that indicates the reliability of the depth/location. This is an interesting project that I think will be useful for other networks looking to incorporate machine-learning algorithms as well as developers of those algorithms who don't always fully understand the constraints of real-time/operational environments. Overall, this manuscript is very well written with clear explanations and very few grammatical errors. The figures are high quality and support the argument. The data and code availability is provided with links. I've provided some suggestions below, mostly minor or optional considerations that could further improve the argument/analysis. This paper will be a nice contribution once revised.

General comments:

What happens if the associator detects more than one event in a window? Does it only keep the one closest to the target event, or does it add the additional events, etc?

Most of the analysis is comparing the depths, since this is the focus; however, it would be interesting to see some comparisons of the epicenters too. For the lower LQS values with accurate depths, perhaps it's the epicenter that's less accurate/more uncertain and lowering the LQS? Given that the covariance matrix determinant was one of the most correlated and highest weighted parameter, it would be reasonable that the LQS is also fairly sensitive to the epicenter uncertainty. It might be interesting to see some plots like fig 4 with the event markers colored by depth accuracy and/or epicenter accuracy (compared to the reference catalogs).

It would be useful to label the referenced place names on at least one map for readers who aren't familiar with this region.

Line comments:

Line 41: Commas should be removed here ("that …" is specifying the type of workflow).

Lines 120-123: It may be worthwhile to include the number of stations in these networks, the types of sensors they use (specifically, broad-band, strong motion, etc.), and an estimate of how many are typically used for a single event. These don't have to be exact (I'm sure they're always changing), just a brief description to get a feel for the types of instruments and network sizes you're working with.

Line 180: The paragraph break here isn't really necessary.

Lines 181-183: The phrasing here is a little awkward, and it's an incomplete sentence. The easiest fix would be changing "improving" to "improves". Also the clause starting with "which" should come after "local seismic noise" since I believe that's what it's referring to?

Lines 197-198: Just to clarify – "event time" is referring to the origin time?

Lines 269-271: Is this adjustment also done in the near real time system (i.e., when it's applied to "new" events)?

Line 283: "criteria aim"

Lines 390-391: It may be helpful to add a little more info about these coefficients for those who aren't knowledgeable about them (could be as simple as value ranges and what they mean).

Lines 394-396: It would be interesting to see plots that more directly compare these (e.g., manual-auto depth difference vs LQS, uncertainty vs LQS).

Line 533: Should this be "manually located events"?

Figure 5: What are the black dots in the center plots? Other tested locations?

Figure 7: Consider adding a note in the caption about the red box zoom-in (I was a little confused at first by the ones on the maps in particular).

Figure 10: Would it be possible to draw lines between the matching events from the two catalogs? It would be interesting to see how they pair up, though it's possible the lines could also make it too messy/cluttered.

Reviewer B:

Review of “A quasi-real-time system for automatic local event monitoring in Germany” by Ramos et al.

The presented procedure to enhance the automatic real-time earthquake monitoring system of the German Federal Seismological Survey and to quantify location reliability is well founded. The location quality score the authors created and the visualisation with the help of radar plots are a helpful tool for analysts to identify and prioritise events with questionable locations for manual intervention. Challenges and shortcomings are identified by the authors and thoroughly addressed. The manuscript is well written and organised and I only have a few comments, questions and suggestions:

It should be mentioned that the criteria for the GT event list that are referenced have very recently been updated and refined (Gallacher et at., 2025; https://doi.org/10.26443/seismica.v4i1.1536). Specifically, the threshold for the secondary azimuthal gap increased to <=210 deg, the requirement of a station within 10km is relaxed if there are sufficient P and S pairs, and a new metric for azimuthal coverage has been introduced. The new criteria have been developed using the ISC’s location algorithm ISCloc and they may not apply to all location techniques. Particularly, replacing the metric for azimuthal coverage relies on the locator’s ability to account for an uneven station distribution. While the current criteria used in this manuscript are suitable for the application, the authors might want to look into the updated criteria to see if some of them would be beneficial for their procedure.

Did you also compare the automatic depths of events with manual fixed depths? I don’t think you mention how many events of the 594 test catalogue were fixed to a depth, but looking at the depth distribution in Figure 2 it seems a substantial portion of the data was fixed to 10 km. I assume the reasons for fixing event depths (e.g., insufficient amount of (S) picks, low coverage, high uncertainties?) would result in a low LQS value.

Minor issues:

Line 82: Adding a sentence providing some information about the grid would be good. I assume that the grid points are more dense in regions of higher seismicity? About how far are the grid points apart in dense and less dense regions? Also, how far are the AD Detector grid points usually away from the final hypocentres?

Figure 1, left: Do you have another figure in a higher resolution? The traces are very blurry and the red and blue lines are hard to see too.

Line 120: Add full name of GFZ

Lines 196 ff: I am not sure I fully understood how you obtain the best/ most probable predicted pick with the multi window approach. You get 4 predicted picks, for each window one, correct? Do you then chose the one with highest probability out of the four or do you group them within a certain time of each other, add them and the timestamp with the highest (grouped) probability is your chosen pick? Or something else?

Line 210 ff: What was used as initial hypocentre for the input of the test data set? The catalogue epicentre or the grid point from the AD Detector? You mention in line 210 '...which in real- time would correspond to detection messages from the AD Detector.' which makes it sound as if for the test you didn't use the detector messages. But in line 220 you say that the accurate preliminary epicentres of the AD Detector allowed to select specific velocity models, which sounds like you did use them as input.

Line 237: S picks play an important role for resolving depths. Would the number of S picks/P and S pairs as a separate parameter be worthwhile or do you usually see a sufficient amount of S phases in the EdB catalogue? Or is this sufficiently covered by the determinant of the covariance matrix?

Line 287 ff: Add something along the line of: LQS of 1 is "good" and 0 is "bad".

Line 300: That looks more like 20 x 20 km, not m. (same in caption for Fig. 4)

Figure 5, caption: 'Left: Event time, epicentre, and associated station distribution' - hypocentre, not epicentre
-The precision of the depth might be a bit too ambitious here (tens of metres range).
-I assume the list of uncertainties in the figure follows the standard x = North, y = East, z = Depth?

Line 390: A brief explanation of when to use the Spearman or Pearson coefficient and what they show would be helpful.

Line 411: 'Notably, events with moderate-to-high LQS values (around 0.5 or greater) cluster along primary azimuthal gaps of approximately 100-120 deg.'- A bit confusing. It is only some events and unless I am miscounting the ticks along the x axis the azimuthal gap in Figure 8 only goes to about 110 degrees.

Line 515: 'A clear agreement trend is observed ....' - While there is a trend in Figure 10a there is also quite a bit of scattering (which you do address in the text). I assume a lot of events with small depth differences are overlain and hidden clustering around 10 – 15 km depth? Maybe showing a histogram of the depth differences in addition would help to give a better idea of the distribution.

Line 518: There are also quite a few events with higher LQS values below 5 km manual depth showing higher depth differences.

Supplementary:
Figure 1: Here, it says there are 591 events shown while the main text claims there are 594.

Response to reviewers

Manuscript:

A quasi-real-time system for automatic local event monitoring in Germany

C. Ramos, S. Donner & K. Stammler

Dear Handling Editor Hongyu Sun,

we thank you and the two reviewers for your time and careful assessment of our original manuscript, which allowed us to submit an improved version.

Overall, we addressed all comments from both reviewers with targeted text clarifications, small methodological notes, figure/caption updates, and two brief supplementary analyses. All newly added text is highlighted (blue) in the revised manuscript; reviewer line references below map to the original submission, while we cite line numbers in the revised version. The core findings and conclusions remain unchanged.

Answers to each reviewers' concern are provided further below and we hope that we have satisfactorily addressed them so that this revision meets the requirements for publication in Seismica.

Sincerely,
C. Ramos, corresponding author on behalf of all co-authors

Reviewer A:

Review of “A quasi-real-time system for automatic local event monitoring in Germany” by Ramos et al.

<u>Recommendation:</u> Revisions Required

Summary:

Study topic/area – Germany, automatic real-time event localization/refinement

Main goal(s) – design an automatic system for refining earthquake hypocenters, particularly depths, and assessing their quality/reliability

Methods/analyses – TieBeNN (PhaseNet picker, PyOcto associator, NonLinLoc), Location Quality Score, catalog comparisons, synthetic events

Results – auto depths generally showed good accuracy with higher uncertainty typically corresponding to lower quality values, LQS wasn’t as meaningful for depths <~30 km

Interpretations/conclusions – TieBeNN has provided benefits to routine monitoring operations, additional adjustments and development could further improve its performance

This manuscript describes the development of a near real-time/post-processing pipeline to refine events detected by the EdB’s existing real-time system, particularly by adding depths, as well as assigning a quality value that indicates the reliability of the depth/location. This is an interesting project that I think will be useful for other networks looking to incorporate machine-learning algorithms as well as developers of those algorithms who don’t always fully understand the constraints of real-time/operational environments. Overall, this manuscript is very well written with clear explanations and very few grammatical errors. The figures are high quality and support the argument. The data and code availability is provided with links. I’ve provided some suggestions below, mostly minor or optional considerations that could further improve the argument/analysis. This paper will be a nice contribution once revised.

General comments:

What happens if the associator detects more than one event in a window? Does it only keep the one closest to the target event, or does it add the additional events, etc?

Thank you for pointing this out. In case the associators used in our workflow detect multiple events within the 60-second time window, we select the associator solution whose origin time is closest to the origin time provided by the AD-Detector message that triggered the TieBeNN run. The P- and S-phases associated with that event are retained, while other phases identified within the same window are discarded for that run. Additional events occurring within the same window are expected to generate separate AD-Detector messages and be processed independently. We have clarified this behavior in the manuscript **(Section 4.2, lines 224–225)**.

Most of the analysis is comparing the depths, since this is the focus; however, it would be interesting to see some comparisons of the epicenters too. For the lower LQS values with accurate depths, perhaps it’s the epicenter that’s less accurate/more uncertain and lowering the LQS? Given that the covariance matrix determinant was one of the most correlated and highest weighted parameter, it would be reasonable that the LQS is also fairly sensitive to the epicenter uncertainty. It might be interesting to see some plots like fig 4 with the event markers colored by depth accuracy and/or epicenter accuracy (compared to the reference catalogs).

We agree that exploring epicenter uncertainty in conjunction with LQS could be informative. To examine its role on LQS more closely, we analyzed all events with non-fixed depths and computed the depth misfit to the reference catalog, $\Delta z = |z_{auto} - z_{manual}|$. We then plotted LQS vs. Δz and LQS vs. depth uncertainty, σ_z, coloring points by the mean horizontal uncertainty, $\bar{\sigma}_h = 0.5 * (\sigma_x + \sigma_y)$:

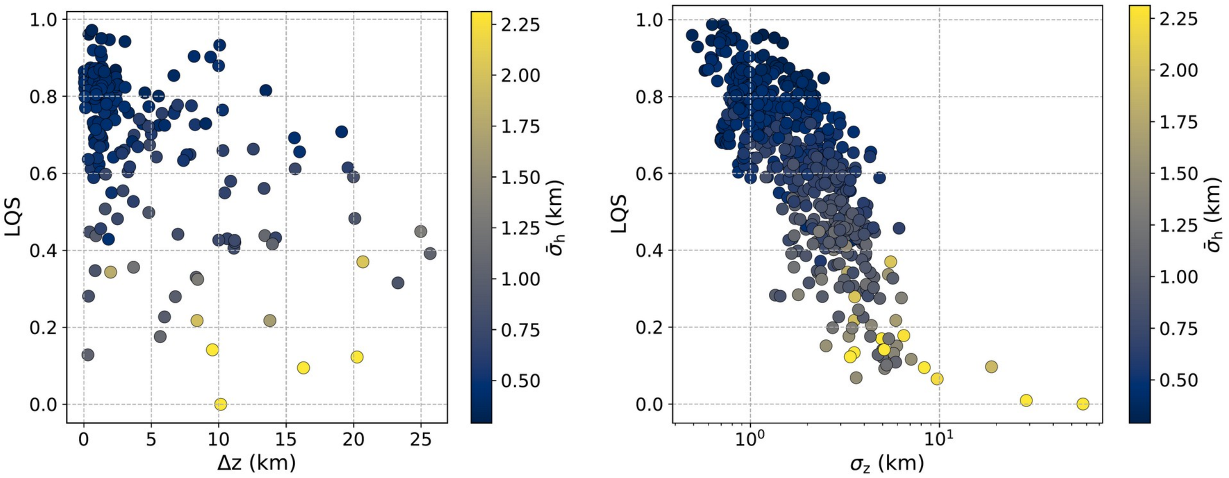

Fig.R1. (left) LQS vs. depth misfit Δz; (right) LQS vs. vertical uncertainty σ_z. Points are colored by the mean horizontal uncertainty $\bar{\sigma}_h$.

The plots show that (i) for high LQS and small $\Delta z/\sigma_z$, $\bar{\sigma}_h$ is typically low; (ii) low-LQS cases at small Δz are uncommon. They tend to exhibit elevated $\bar{\sigma}_h$, consistent with the reviewer's hypothesis that weaker epicentral constraint contributes to lower LQS. However, in our dataset, sparse station geometry often causes horizontal and vertical uncertainties to increase together, so LQS, through the covariance determinant, is primarily sensitive to their joint inflation rather than cleanly separating horizontal from vertical contributions.

We also note that depth misfit to reference catalogs is not an absolute accuracy metric (see discussion in Section 6.1): event depths can differ substantially among agency catalogs (see e.g. Fig.10b) due to e.g. the selection of distinct velocity models for event location, explaining some high LQS points with large Δz despite small $\bar{\sigma}_h$.

We included Fig.R1 as **Supplementary Figure S5** and added a brief paragraph in the Discussion **(Section 6.1, lines 541–548)**.

It would be useful to label the referenced place names on at least one map for readers who aren't familiar with this region.

We added country/region labels and key place names across **Figures 2, 5–7**, and clarified the location of Rhineland-Palatinate in the text **(Sec. 4.4, lines 319–320)**. Figure captions were updated to define some abbreviations.

Line comments:

Line 41: Commas should be removed here ("that …" is specifying the type of workflow).

Comma was deleted.

Lines 120-123: It may be worthwhile to include the number of stations in these networks, the types of sensors they use (specifically, broad-band, strong motion, etc.), and an estimate of how many are typically used for a single event. These don't have to be exact (I'm sure they're always changing), just a brief description to get a feel for the types of instruments and network sizes you're working with.

We added approximate network sizes in **Section 3.1, lines 122–127**. The instrument types (broadband and short-period) were already stated in **Section 3.1 lines 120–121**.

Line 180: The paragraph break here isn't really necessary.

Paragraphs have been merged **(Section 3.4, line 184)**.

Lines 181-183: The phrasing here is a little awkward, and it's an incomplete sentence. The easiest fix would be changing "improving" to "improves". Also the clause starting with "which" should come after "local seismic noise" since I believe that's what it's referring to?

We rephrased this sentence for clarity and to fix the fragment, moving the 'which' clause next to 'local seismic noise' **(Section 3.4, lines 185–187)**.

Lines 197-198: Just to clarify – "event time" is referring to the origin time?

Yes. We now state this explicitly **(Section 3.4, lines 202–203)**.

Lines 269-271: Is this adjustment also done in the near real time system (i.e., when it's applied to "new" events)?

Yes. The "pre-normalization guard" against spuriously low RMS arising from very few picks is applied operationally in the near-real-time system in the same way as in our test. The thresholds on RMS and station density were derived from our 594-event catalog and are currently fixed. As the system gathers more automatic locations, we will be able to periodically reassess these thresholds, adapting them if needed (including region-specific tuning) to ensure that the RMS normalization continues to suppress this artifact without masking genuinely low-RMS, well-constrained solutions. We have clarified this in the manuscript **(Section 4.3, lines 283–286)**.

Line 283: "criteria aim"

It has been corrected **(Section 4.3, line 297)**.

Lines 390-391: It may be helpful to add a little more info about these coefficients for those who aren't knowledgeable about them (could be as simple as value ranges and what they mean).

We added a brief explanation of what both coefficients mean and why we report both **(Section 5.2, lines 408–414)**.

Lines 394-396: It would be interesting to see plots that more directly compare these (e.g., manual-auto depth difference vs LQS, uncertainty vs LQS).

We agree that direct comparisons aid interpretation. The plots suggested by the reviewer were provided in **Supplementary Figure S5** (also shown as Response Fig. R1) as part of our analysis of horizontal-uncertainty and depth-accuracy effects on LQS.

Line 533: Should this be “manually located events”?

Thank you for checking. In this case **(now Section 6.2, line 569)**, “automatically located events” is correct. The LQS-parameter normalization was derived from the 594 events automatically located by our TieBeNN workflow (NonLinLoc solutions) in our test. Manual catalogs were used only to trigger TieBeNN runs: for each event they provided latitude, longitude, and origin time to simulate the AD-Detector message that TieBeNN “listens to” in quasi-real time.

Figure 5: What are the black dots in the center plots? Other tested locations?

The black dots in the center panels depict grid samples with non-negligible posterior probability from NonLinLoc’s probabilistic location; they illustrate the uncertainty cloud rather than additional tested solutions. We clarified this in the Figure 5 caption.

Figure 7: Consider adding a note in the caption about the red box zoom-in (I was a little confused at first by the ones on the maps in particular).

We added a brief note to the Figure 7 caption clarifying this.

Figure 10: Would it be possible to draw lines between the matching events from the two catalogs? It would be interesting to see how they pair up, though it’s possible the lines could also make it too messy/cluttered.

Connecting segments would make the figure indeed too cluttered. Instead, we clarified the caption in Figure 10 to explain how pairs can be identified: automatic depth (y-axis) and LQS color are identical in both symbols.

Reviewer B:

Review of "A quasi-real-time system for automatic local event monitoring in Germany" by Ramos et al.

<u>Recommendation:</u> Revisions Required

The presented procedure to enhance the automatic real-time earthquake monitoring system of the German Federal Seismological Survey and to quantify location reliability is well founded. The location quality score the authors created and the visualisation with the help of radar plots are a helpful tool for analysts to identify and prioritise events with questionable locations for manual intervention. Challenges and shortcomings are identified by the authors and thoroughly addressed. The manuscript is well written and organised and I only have a few comments, questions and suggestions:

It should be mentioned that the criteria for the GT event list that are referenced have very recently been updated and refined (Gallacher et at., 2025; https://doi.org/10.26443/seismica.v4i1.1536). Specifically, the threshold for the secondary azimuthal gap increased to <=210 deg, the requirement of a station within 10km is relaxed if there are sufficient P and S pairs, and a new metric for azimuthal coverage has been introduced. The new criteria have been developed using the ISC's location algorithm ISCloc and they may not apply to all location techniques. Particularly, replacing the metric for azimuthal coverage relies on the locator's ability to account for an uneven station distribution. While the current criteria used in this manuscript are suitable for the application, the authors might want to look into the updated criteria to see if some of them would be beneficial for their procedure.

Thank you for this timely pointer. Our original submission predates Gallacher *et al*. (2025). We now acknowledge the updated GT criteria in the Discussion, noting (a) the relaxed near-station rule when sufficient P–S pairs exist, and (b) the introduction of the Cyclic Polygon Quotient (CPQ) for azimuthal coverage. As our LQS already emphasizes station density/uniformity over nearest-station distance, it is broadly aligned with the spirit of the update; and given the comparatively weak correlation of azimuthal gap with LQS in our dataset, we agree that CPQ is a promising replacement to evaluate in future work. The new text has been added in **Section 6.2, lines 580–586**.

Did you also compare the automatic depths of events with manual fixed depths? I don't think you mention how many events of the 594 test catalogue were fixed to a depth, but looking at the depth distribution in Figure 2 it seems a substantial portion of the data was fixed to 10 km. I assume the reasons for fixing event depths (e.g., insufficient amount of (S) picks, low coverage, high uncertainties?) would result in a low LQS value.

We did analyze the fixed-depth cases but omitted them from the original draft. We now (i) state the fixed/non-fixed counts in the catalog in **Section 4.1 lines 212–213**, (ii) add **Supplementary Figure S6** showing the automatic depths versus all manual depths for completeness, and (iii) include a brief note in **Section 6.1, lines 558–563** explaining the expected clustering at some fixed depths and the tendency toward lower LQS (with a small subset attaining high LQS where coverage improved in the automatic location).

Minor issues:

Line 82: Adding a sentence providing some information about the grid would be good. I assume that the grid points are more dense in regions of higher seismicity? About how far are the grid points apart in dense and less dense regions? Also, how far are the AD Detector grid points usually away from the final hypocentres?

We added a sentence describing the grid density and typical spacings, and clarified proximity to final source solutions. The irregular grid's density follows regional seismicity. In Germany, spacing is ~20–40 km, and preliminary epicenters are typically within <10 km of final solutions. These clarifications are now included in the text **(Section 2.2, lines 84–85 and 94–95)**.

Figure 1, left: Do you have another figure in a higher resolution? The traces are very blurry and the red and blue lines are hard to see too.

Figure 1 depicts an operational quick-look graphic automatically produced by the AD-Detector for analyst orientation rather than detailed waveform inspection, which is carried-out separately using a different software. We now clarify this purpose in the caption.

Line 120: Add full name of GFZ

We expanded the acronym at first mention **(Section 3.1, line 123)**.

Lines 196 ff: I am not sure I fully understood how you obtain the best/ most probable predicted pick with the multi window approach. You get 4 predicted picks, for each window one, correct? Do you then chose the one with highest probability out of the four or do you group them within a certain time of each other, add them and the timestamp with the highest (grouped) probability is your chosen pick? Or something else?

Each of the four window alignments yields at most one P and/or S prediction in the end. We retain, for each phase independently, the timestamp with the highest predicted probability across the four windows. We clarified the selection criterium in **Section 3.4, lines 204–205**.

Line 210 ff: What was used as initial hypocentre for the input of the test data set? The catalogue epicentre or the grid point from the AD Detector? You mention in line 210 '...which in real- time would correspond to detection messages from the AD Detector.' which makes it sound as if for the test you didn't use the detector messages. But in line 220 you say that the accurate preliminary epicentres of the AD Detector allowed to select specific velocity models, which sounds like you did use them as input.

We have clarified the wording. For the test dataset, the initial hypocenter input was the manual-catalog epicenter (not the AD-Detector grid point). The references to the AD-Detector were forward-looking: in real-time operation, TieBeNN will (and now does) listen to the AD-Detector's detection messages, using the preliminary epicenter to drive the same automatic velocity-model selection. We revised the text accordingly to avoid ambiguity **(Section 4.2, lines 217–219 and 230–233)**.

Line 237: S picks play an important role for resolving depths. Would the number of S picks/P and S pairs as a separate parameter be worthwhile or do you usually see a sufficient amount of S phases in the EdB catalogue? Or is this sufficiently covered by the determinant of the covariance matrix?

In our setting (GRSN + neighboring networks; PhaseNet/EQTransformer for automatic phase pick), we typically observe ample S-phase coverage, so the covariance determinant already captures much of the P/S constraint through the resulting uncertainty ellipsoid. That said, we agree that an explicit measure could be useful. One idea to test in the future could be replacing the "number of picks" parameter with the "number of stations carrying both P and S picks", which would emphasize depth-resolving power and likely reduce redundancy with station density. Since total picks already correlate strongly with LQS (see Figure 8), we expect the "stations with P–S pairs" variant to be at least as informative, and possibly more so for depth. Implementing and re-running the full evaluation is out of scope for this revision, but we have added a note in **Section 6.2, lines 586–590** committing to test this parameter alongside CPQ (Gallacher *et al.*, 2025) in future work.

Line 287 ff: Add something along the line of: LQS of 1 is “good” and 0 is “bad”.

We added a brief statement clarifying the interpretation of the scale **(Section 4.3, line 306)**.

Line 300: That looks more like 20 x 20 km, not m. (same in caption for Fig. 4)

Thank you for catching this. This error on our side originated in GMT/PyGMT using "m" to denote a grid spacing of arc-minutes, not meters. We mistakenly wrote "20×20 m." We have corrected the text **(Section 4.4, line 315)** and Figure 4 caption to "20'×20'".

Figure 5, caption: ‘Left: Event time, epicentre, and associated station distribution’ - hypocentre, not epicentre
-The precision of the depth might be a bit too ambitious here (tens of metres range).
-I assume the list of uncertainties in the figure follows the standard x = North, y = East, z = Depth?

- We corrected "epicenter" to "hypocenter" in the Figure 5 caption.

- We reduced the displayed depth precision in Figure 5 to match the reported uncertainty, now shown to one decimal place.

- We clarified the axis convention in the caption as x = East, y = North, z = Depth, consistent with NonLinLoc's output.

In our original draft, some figures used NonLinLoc’s confidence-ellipsoid semi-axes in place of component uncertainties. For clarity and consistency we now report component-wise 1-D standard deviations from the covariance matrix: $\sigma_x = \sqrt{C_{xx}}$, $\sigma_y = \sqrt{C_{yy}}$, $\sigma_z = \sqrt{C_{zz}}$ (x = East, y = North, z = Depth). We updated the location catalogs accordingly (see the revised Data and code availability section, **line 686**) and regenerated the affected figures:

- Figure 5 (center): numeric uncertainty labels now reflect σ_x, σ_y, σ_z.
- Figures 6–7: depth-uncertainty error bars in scatter plots updated.
- Figure 9a: curves unchanged; only the y-axis range reflects the revised σ_z.

Likewise, we revised the manuscript text to reflect the corrected uncertainty values in **Section 4.5 (line 339)**, **Section 5.1 (line 377)**, and **Section 5.2 (line 395)**.

Please note that this modification does not alter our conclusions. The LQS determinant term in our analysis has always been computed from the covariance matrix **C**, and NonLinLoc’s ellipsoid semi-axes are related to **C**’s eigenvalues by a constant coverage factor (for 68% in 3-D, is approximately 1.879). Thus the qualitative trends and parameter rankings remain the same; only the displayed magnitudes of uncertainty values have been corrected to the component standard deviations.

Line 390: A brief explanation of when to use the Spearman or Pearson coefficient and what they show would be helpful.

This was also a concern of Reviewer A and we added a brief explanation of both coefficients **(Section 5.2, lines 408–414)**.

Line 411: ‘Notably, events with moderate-to-high LQS values (around 0.5 or greater) cluster along primary azimuthal gaps of approximately 100-120 deg.’- A bit confusing. It is only some events and unless I am miscounting the ticks along the x axis the azimuthal gap in Figure 8 only goes to about 110 degrees.

Thank you, the phrasing was ambiguous. We did not mean that all moderate-to-high LQS events cluster at large primary gaps; rather, a subset does. We verified the underlying data and actually found two points between 120° and 130°. Accordingly, we revised the text. See **Section 5.2, lines 433–436**.

Line 515: ‘A clear agreement trend is observed ....’ - While there is a trend in Figure 10a there is also quite a bit of scattering (which you do address in the text). I assume a lot of events with small depth differences are overlain and hidden clustering around 10 – 15 km depth? Maybe showing a histogram of the depth differences in addition would help to give a better idea of the distribution.

Thank you for the suggestion. To complement Figure 10a and avoid possible overplotting near small misfits, we added a histogram of signed depth differences as **Supplementary Figure S4**, which clarifies the distribution and any bias. We reference this addition in **Section 6.1, lines 538–539**.

Line 518: There are also quite a few events with higher LQS values below 5 km manual depth showing higher depth differences.

We agree. We revised the wording to encompass shallow manual depths below 5 km as well as 5–10 km **(Section 6.1, line 549)**.

Supplementary:

Figure 1: Here, it says there are 591 events shown while the main text claims there are 594.

Thank you for catching this typo in the Supplement. We corrected it.

Round 2

Reviewer A:

Review of revision of “A quasi-real-time system for automatic local event monitoring in Germany” by Ramos et al.

Summary:

I’d like to thank the authors for their work in revising this manuscript. My concerns have been satisfactorily addressed along with the other reviewer’s. I’ve included a few minor edits below. Congrats on the great paper!

Line comments (annotated version):

Line 41: remaining comma should also be removed

Line 226: “a few”

Line 413: remove period after "complementary"

Lines 580-590: I’d suggest moving this new paragraph after line 599. That way, you’re continuing to discuss the LQS results first, before making suggestions for future improvements.

Supp Fig S6: Please state what the x and o markers indicate (I assume fixed or not fixed depth).

Reviewer B:

The authors have carefully addressed all comments and I recommend to accept the submission.

I have just one note: My comment 'Add something along the line of: LQS of 1 is “good” and 0 is “bad” didn't mean to literally add: 1 = “good”, 0 = “bad”, but to explain what 1 and 0 stand for in general. Which the authors did. Feel free to remove the last part of the sentence in line 306 (tracked) in parentheses, if you wish.