

Round 1

Reviewer A:

For author and editor

This manuscript presents and compares two workflows for detecting and locating earthquakes. Both workflows use PhaseNet for phase picking. The first workflow PRN follows a conventional procedure of locating earthquakes: phase picking, association and location. The second workflow locates earthquakes by backprojecting the phase annotation of the deep learning model. The authors have provided a very detailed description of their workflow. Although PhaseNet was originally trained on tectonic earthquakes, the authors demonstrate that their workflow can work well for a region with reservoir impoundment and hydrocarbon extraction, achieving a recall up to 93%, though false positives and precisions are difficult to quantify. Overall, the manuscript is clearly written. Please find my comments below.

Line 56: Qseek appears here for the first time. It is one of the two methods used in the paper, so I think a brief introduction about Qseek is necessary.

Lines 58-59: Could you briefly describe how operators revise the catalog? For example, do they manually check and repick the phase arrival by zooming in the waveforms, or just quickly glance to check whether there is a signal?

Line 81: The PhaseNet variant trained on an Italian dataset INSTANCE may be more suitable for your data.

Line 103: How do you find the peak-to-peak amplitude?

Lines 113-114: It seems that the purpose of stacking is to calculate the coherence of different stations instead of enhancing the SNR.

Line 118: What is the threshold of semblance used in this study?

Line 119: “the semblance value is positively correlated with earthquake magnitude.” Do you have any references or evidence to support this statement? As far as I know, the semblance is calculated by stacking the pick confidence curves output by the deep learning picker at different stations, so it is positively correlated with pick confidence (or probability). This statement might be true in some cases because models tend to give higher pick confidence to waveforms with higher SNR and larger earthquakes usually have higher SNR. However, pick confidence also depends on how well the tested waveform can be represented by the distribution of the training data set. For out-of-distribution data, such as large earthquakes, low-frequency earthquakes, the pick confidence might be low even if the magnitude is large. It can be misleading to say semblance is correlated with magnitude if the readers don’t know much detail of the semblance in Qseek.

Line 121: Do you consider only stations within 20 km of the epicenter for magnitude calculation only in the Qseek workflow?

Line 122: How do you know the magnitude is overestimated if you don't know the real magnitude beforehand?

Line 133: Alternatively, you can randomly select some examples (e.g. 100 examples) from each of your catalogs and visually count the number of true and false positives. In this way, you can have a rough estimate of the precisions of the two workflows.

Line 134, line 161: How do you compare events from two catalogs? Do you compare their event location and origin time? What are the thresholds of the differences?

Figure 1b: True positive rate is recall. It seems that you are showing the precision.

Line 187: When is the change?

Line 188: I find the sentence a bit mouthful.

Figures 7-8: Could you please show all the three catalogs?

Line 222: "... preserving similarity between the two distributions at higher magnitudes". Figure 2 shows all the detected events. To see the similarity clearly, it may be helpful to add a figure of only the events above the M_c of the CMS catalog in supplementary information.

Line 263: Could you show the lake and the selected region on the map?

Discussion: there are many aspects in Discussion, the authors can consider grouping them into subsections.

Reviewer B:

Dear Editor,

I have reviewed the manuscript entitled “*Improving detection of micro-earthquakes in the Val d’Agri region (Southern Italy) using deep learning algorithms*” with great interest. The authors present a compelling study that evaluates the application of two deep-learning-based workflows, using advanced algorithms such as PhaseNet, REAL, NonLinLoc, and Qseek, to detect and locate small-magnitude earthquakes in the Val d’Agri region, Southern Italy.

The manuscript shows that modern machine learning tools can outperform classical detection techniques (e.g., STA/LTA, SNR-based methods), especially in improving sensitivity for the detection of microseismic events. The resulting catalogs detect at least twice as many events compared to the conventional workflow, which is a substantial step forward for automatic seismic monitoring.

The study is conceptually solid, and the manuscript is generally well structured. The methods fit the study goals, and the results are supported by quantitative analysis. Still, some parts, especially those describing the workflows, data processing steps, and selection thresholds, could be clarified to make the paper more accessible for readers not very familiar with these techniques. I also noticed a lack of key references in this field; adding some of them could help the introduction and further discussion. The discussion could also better highlight the potential impact for real-time monitoring or use in other industrial settings.

Improving these aspects would make the paper clearer and help others reproduce the work.

Overall, I think this contribution fits the scope of the journal and I would recommend it for publication after minor revision.

Minor comments and suggestions

L36–L38: Consider citing more recent works that reflects the current state of deep learning in seismic detection and catalog creation. This would help place the study in a broader context.

L78–L80: I guess the 1–45 Hz bandpass filter is based on the Nyquist frequency, but could you also comment on whether it reflects specific characteristics of the events or the local noise? That would help with reproducibility.

L81–L82: Why was the PhaseNet pick score threshold set to 0.3? Was this value based on previous studies, or is a specific value for this dataset? Maybe some cites would be useful to support it.

L83–L84: The choice of a minimum of 7 phases is not clearly explained. Can you give a bit more background or reasoning for this?

L84–L85: Is the 2 second residual cutoff in REAL a default value?, or did you choose it based on your results? A bit more explanation here would help readers follow your workflow.

L100–Eq. 1: I believe the parameters in this equation are standard for California, right? It probably works well enough, but can you add some support or justification for using it here as a proxy for ML?

L105: The CMS catalog also uses the same equation and parameters for ML? I think you mention this later, but it might help to say it earlier for clarity.

L133–L150: You justify the filtering using semblance–magnitude and average station distance, which is good. But I’d suggest adding a few waveform examples in the Supplementary, showing low-magnitude events that passed the filters. That would help readers evaluate the quality.

L161: The phrase “To compare catalog, we compare...” should be rephrased to improve the reading flow.

L167–L168: What do you mean by “high quality from NonLinLoc”? Can you add some quantitative explanation of what makes these locations better?

L183–L185: This paragraph could be rephrased for clarity.

L197: When you say “attention zone,” do you mean the whole study area? Maybe use clearer wording here.

L201: You mentioned uncertainties from NonLinLoc earlier. Could you discuss general location uncertainties a bit more? That would help understand the quality of the catalogs' location.

L239: Typo: “antropogenic” should be “anthropogenic”.

L270–L274: This paragraph is a bit unclear. The phrase “slight increase in seismicity” followed by “number of earthquakes remains low” is confusing. Try rephrasing and tie it more directly to what’s shown in Figure 8.

Figures

Figure 2: The color code by depth works well in the plain map view, but not in the cross-sections (since depth is already on the y-axis). You could use a plain color or color by something else in the cross-section. Also, I think the circle size reflects magnitude, if so, a legend mentioning that is needed. A colorbar is also missing for the plain view. Maybe add the number of events shown too? Also, the x-axis label is missing in the cross-sections; something like “Distance along profile” would work.

Figure 3: Please add a colorbar and the same point mentioned for Figure 2.

Response to reviewers

Improving detection of micro-earthquakes in the Val d’Agri region (Southern Italy) using deep learning algorithms

Caredda E., Isken M.P., Cesca S., Errico M., Zerbinato G., Morelli A.

July 21, 2025

We would like to sincerely thank the Editor and the Reviewers for their thoughtful and constructive comments, that have helped us improve the clarity and quality of our manuscript. We appreciate the time and effort they dedicated to reading and evaluating our work. We have revised the manuscript following the recommendations. We provide a detailed response to each comment, copied here with underlined font, with reference to line numbers in the original manuscript. Where we indicate the changes we reference instead the relevant line numbers in the revised version.

1 Reviewer A

This manuscript presents and compares two workflows for detecting and locating earthquakes. Both workflows use PhaseNet for phase picking. The first workflow PRN follows a conventional procedure of locating earthquakes: phase picking, association and location. The second workflow locates earthquakes by backprojecting the phase annotation of the deep learning model. The authors have provided a very detailed description of their workflow. Although PhaseNet was originally trained on tectonic earthquakes, the authors demonstrate that their workflow can work well for a region with reservoir impoundment and hydrocarbon extraction, achieving a recall up to 93%, though false positives and precisions are difficult to quantify. Overall, the manuscript is clearly written. Please find my comments below.

Line 56: Qseek appears here for the first time. It is one of the two methods used in the paper, so I think a brief introduction about Qseek is necessary.

Thanks for pointing out the need for more detail. We added the following paragraph to introduce

Qseek (at lines 67-69 in the revised manuscript):

Qseek (Isken et al., 2025) is a data-driven method that combines phase arrival annotations from pre-trained neural networks with waveform stacking and an adaptive octree search algorithm to detect and locate earthquakes automatically.

Lines 58-59: Could you briefly describe how operators revise the catalog? For example, do they manually check and repick the phase arrival by zooming in the waveforms, or just quickly glance to check whether there is a signal?

Indeed, this is a relevant piece of information. The operators manually check every detection and repick the phase arrivals when necessary. We have now clarified this point in the text (line 71).

Line 81: The PhaseNet variant trained on an Italian dataset INSTANCE may be more suitable for your data.

We thank the reviewer for the suggestion. The PhaseNet variant trained on the INSTANCE dataset (Michellini et al., 2021) — composed of seismic data from across Italy — could indeed be considered better adapted to our region than the original model trained on Californian data. However, a recent study by (Cianetti et al., 2025), conducted in the Central Apennines, directly compares the performance of the original PhaseNet model and the INSTANCE-trained variant. The results show only minor differences in picking quality and event detection. Given the similar tectonic setting and waveform characteristics between the Central and Southern Apennines, we expect comparable performance from the original PhaseNet model in our case as well. Therefore we think that our choice is well justified. We have added a brief comment and the relevant reference to the manuscript to clarify this point (see lines 102–105).

Line 103: How do you find the peak-to-peak amplitude?

The peak-to-peak amplitude is calculated within a window that starts 0.5 seconds before the P-wave arrival and ends 3 seconds after the expected S-wave arrival time. The peak-to-peak value is computed using the `numpy.ptp` function on this time window. We have expanded the explanation of the magnitude calculation in the revised manuscript (lines 139–140).

Lines 113-114: It seems that the purpose of stacking is to calculate the coherence of different stations instead of enhancing the SNR.

Indeed, we agree with the Reviewer that the primary purpose of stacking in this context is to assess the coherence of phase picks across multiple stations. However, we note that this process inherently enhances the signal-to-noise ratio (SNR) as well. In particular, coherent peaks corresponding to

true seismic arrivals are reinforced through stacking, while incoherent noise, which is randomly distributed across stations, is attenuated. This results in higher effective SNR in the stacked signal. We clarified this aspect in the revised manuscript (lines 149–152).

Line 118: What is the threshold of semblance used in this study?

We thank the Reviewer for this question, that prompts us to better describe our study. We apply an adaptive threshold to the semblance trace, based on the Median Absolute Deviation (MAD) — a robust measure of variability, calculated from the maximum semblance trace — to ensure the detection of low-magnitude earthquakes. We added a short explanatory paragraph in the revised manuscript (lines 157–159).

Line 119: “The semblance value is positively correlated with earthquake magnitude.” Do you have any references or evidence to support this statement? As far as I know, the semblance is calculated by stacking the pick confidence curves output by the deep learning picker at different stations, so it is positively correlated with pick confidence (or probability). This statement might be true in some cases because models tend to give higher pick confidence to waveforms with higher SNR and larger earthquakes usually have higher SNR. However, pick confidence also depends on how well the tested waveform can be represented by the distribution of the training data set. For out-of-distribution data, such as large earthquakes, low-frequency earthquakes, the pick confidence might be low even if the magnitude is large. It can be misleading to say semblance is correlated with magnitude if the readers don’t know much detail of the semblance in Qseek.

Thank you for your valuable comment. We agree that the semblance value is fundamentally based on the coherence of the pick confidence curves across stations, as you pointed out. The semblance is influenced not only by pick confidence but also by the overall data quality and signal-to-noise ratio (SNR).

For very small events, especially those located farther from the stations, the SNR tends to be low at many stations. This results in lower Phasenet pick confidence values and often no picks at distant stations, which in turn leads to a low semblance value. Conversely, for larger events, the seismic signal is typically recorded clearly by most stations in the network, contributing coherently to the semblance and yielding higher semblance values.

Therefore, while the semblance is not a direct measure of magnitude, it is indirectly correlated with it through the spatial extent and quality of the seismic signal recorded across the network. The revised text now reads (lines 159–163 in the revised manuscript):

Although semblance is primarily a measure of the coherence of pick confidence across stations, it is also influenced by the signal-to-noise ratio (SNR) and data quality. For

smaller events, especially those distant from the network, low SNR and fewer picks at some stations tend to reduce semblance values. Conversely, larger events generally produce clearer signals at most stations, resulting in higher semblance values. Therefore, semblance is often indirectly correlated with earthquake magnitude through these effects.

Line 121: Do you consider only stations within 20 km of the epicenter for magnitude calculation only in the Qseek workflow?

Yes, that is correct! We have adopted this distance threshold specifically for the Qseek workflow, because by default it includes all available stations with SNR higher than some threshold (2 in our case) in the time window of the theoretical arrival, even when the station is not triggered. For micro-earthquakes, distant stations often record only noise, leading to unreliable amplitude measurements and, consequently, to an overestimation of the magnitude. We visually inspected several waveforms from distant stations and confirmed that signal quality was generally poor. In fact, in our dataset, many noisy stations pass this threshold, but looking at the waveforms, we find that this is because of some random spikes in the recordings, which are not related to an earthquake signal. Therefore, to ensure consistent and reliable magnitude estimation, we limited the calculation to stations within 20 km from the epicenter. The same cutoff is not necessary for the other workflow, since it only considers the triggered stations. This clarification has been added to the revised manuscript (lines 172-173).

Line 122: How do you know the magnitude is overestimated if you don't know the real magnitude beforehand?

Indeed, this requires some explanation, we thank the Reviewer for pointing out. We use as reference magnitude values those from our reference CMS catalog. We did a comparison on the subset of common events between the two catalogs. Hence, we say it is overestimated compared to the magnitude of the same event in the CMS catalog. We have clarified this point in the manuscript at line 169.

Line 133: Alternatively, you can randomly select some examples (e.g. 100 examples) from each of your catalogs and visually count the number of true and false positives. In this way, you can have a rough estimate of the precisions of the two workflows.

Thanks for your suggestion. We performed the manual visual inspection of a random subset of detections for the Qseek-based workflow. This choice was motivated by the nature of our workflow, which is characterized by a very large number of low-SNR detections due to intentionally permissive detection thresholds (to maximize the recall). As a result, a significant fraction of false positives was expected, making a rough estimate of precision particularly useful and necessary to assess

the effect of subsequent filtering steps (e.g., magnitude–semblance). We added the results at lines 181-187.

Line 134, line 161: How do you compare events from two catalogs? Do you compare their event location and origin time? What are the thresholds of the differences?

Correct, we compare events as identified on the basis of both origin time and epicentral proximity. We define a match if the origin time difference is less than 1.5 seconds and the epicentral distance is less than 10 km. These thresholds have been chosen to account for the typical uncertainties in both parameters for small local earthquakes. We added this detail to the revised manuscript (lines 189-190).

Figure 1b: True positive rate is recall. It seems that you are showing the precision.

Yes, this is correct — the figure shows the number of true positives over the total number of detections in each bin, which corresponds to precision rather than recall. We have updated the figure title to reflect this and avoid confusion.

Line 187: When is the change?

Thank you for your comment. The change in operational procedures that improved the detection threshold in the CMS catalog occurred on May 18, 2023. We have now explicitly added this date in the revised manuscript to clarify the timing of this change (line 255).

Line 188: I find the sentence a bit mouthful.

Done, we rephrased the text (lines 251-254):

Figure 5 shows that the rise in the number of events observed in 2023, along with larger magnitudes, reflects a general increase in seismic activity consistently captured by all three catalogs.

Figures 7-8: Could you please show all the three catalogs?

We thank the Reviewer for the suggestion, and the interest to see additional information, that we are happy to provide. Plotting 3 catalogs on the same figure would worsen readability of Figures 7 and 8, so we prefer to plot results for the other catalogs in additional figures, that we have added as Supplementary Material (Figures S6-S7).

Line 222: "... preserving similarity between the two distributions at higher magnitudes". Figure 2 shows all the detected events. To see the similarity clearly, it may be helpful to add a figure of only the events above the M_c of the CMS catalog in supplementary information.

Following the Reviewer’s suggestion, we added these maps at Figure S5 in the Supplementary. To provide a more robust magnitude threshold for comparing the two catalogs, we estimated the detection probability function using the Ogata Katsura (1993) model. From this, we derived the magnitude corresponding to a 99% probability of detection (M99), which we used to filter the events in the supplementary figure. This approach provides a more objective and statistically sound estimate of the magnitude of completeness than commonly used empirical methods such as the Maximum Curvature, which tend to underestimate the completeness level (e.g., Woessner Wiemer, 2005).

Based on this model, we also obtained a more reliable estimate of the b-value, which accounts for detection biases in the lower-magnitude range. This updated b-value was used to replace the previous estimate in Table 2.

Line 263: Could you show the lake and the selected region on the map?

We have updated Figure 2a to include an inset highlighting the selected region, which also contains the Pertusillo Lake, in order to provide a clearer spatial context.

Discussion: there are many aspects in Discussion, the authors can consider grouping them into subsections.

We have updated the Discussion section into subsections “Comparison of the workflows performance”, “Noise and performance of the detection”, and “Pertusillo water level and seismicity rate”.

2 Reviewer B

Dear Editor,

I have reviewed the manuscript entitled “Improving detection of micro-earthquakes in the Val d’Agri region (Southern Italy) using deep learning algorithms” with great interest. The authors present a compelling study that evaluates the application of two deep-learning-based workflows, using advanced algorithms such as PhaseNet, REAL, NonLinLoc, and Qseek, to detect and locate small-magnitude earthquakes in the Val d’Agri region, Southern Italy. The manuscript shows that modern machine learning tools can outperform classical detection techniques (e.g., STA/LTA, SNR-based methods), especially in improving sensitivity for the detection of microseismic events. The resulting catalogs detect at least twice as many events compared to the conventional workflow, which is a substantial step forward for automatic seismic monitoring. The study is conceptually solid, and the manuscript is generally well structured. The methods fit the study goals, and the results are supported by quantitative analysis. Still, some parts, especially those describing the workflows, data processing steps, and selection thresholds, could be clarified to make the paper more accessible for readers not very familiar with these techniques. I also noticed a lack of key references in this field; adding some of them could help the introduction and further discussion. The discussion could also better highlight the potential impact for real-time monitoring or use in other industrial settings. Improving these aspects would make the paper clearer and help others reproduce the work. Overall, I think this contribution fits the scope of the journal and I would recommend it for publication after minor revision.

Minor comments and suggestions

L36–L38: Consider citing more recent works that reflects the current state of deep learning in seismic detection and catalog creation. This would help place the study in a broader context.

We thank the Reviewer for the very useful suggestion. We have added some references to more recent work (lines 37-48):

Recent advances in deep learning have significantly expanded the range of tools available for seismic monitoring. For instance, EQTransformer employs an attention-based multitask architecture for joint phase picking and event detection, leading to substantial improvements in catalog completeness (Mousavi et al., 2020). Similarly, QuakeFlow integrates PhaseNet and GaMMA into a scalable, cloud-based workflow for efficient earthquake cataloging (Zhu et al., 2022a; Zhu and Beroza, 2019; Zhu et al., 2022). The development of frameworks such as SeisBench further supports the standardization and benchmarking of machine learning models, including PhaseNet and EQTransformer

(Woollam et al., 2022).

In addition to event detection and picking, Bayesian deep learning approaches have been proposed for estimating earthquake source parameters and their uncertainties from single-station recordings (Mousavi and Beroza, 2020). A recent review by Mousavi and Beroza (2023) underscores the wide applicability of machine learning across the entire seismic workflow—from detection and picking to event clustering, focal mechanism analysis, and ground-motion prediction—while highlighting the importance of open data and standardized evaluation practices.

L78–L80: I guess the 1–45 Hz bandpass filter is based on the Nyquist frequency, but could you also comment on whether it reflects specific characteristics of the events or the local noise? That would help with reproducibility.

We agree, and we gladly follow the useful suggestion. For our local network configuration and for the purpose of detecting micro-earthquakes, we select a broad frequency range (1–45 Hz) in order to retain as much signal content as possible. Micro-earthquakes radiate energy at higher frequency, and we found that including frequency up to 45 Hz improves their detectability. As the Reviewer remarks, the upper bound is limited by the Nyquist frequency, as our data are sampled at 100 Hz. Moreover, the 1–45 Hz band is commonly used in similar studies (e.g., Mousavi et al., 2020; Wickham-Piotrowski et al., 2023), which supports its general applicability. We have added a clarification and new citations in the revised manuscript (lines 92-97).

L81–L82: Why was the PhaseNet pick score threshold set to 0.3? Was this value based on previous studies, or is a specific value for this dataset? Maybe some cites would be useful to support it.

We concur with the Reviewer that this explanation is due, and we thank them for their attentiveness. We set the threshold to 0.3 on the PhaseNet pick score to filter out low-confidence picks. This value was chosen based on empirical tests on our dataset and is consistent with thresholds adopted in previous studies using PhaseNet (e.g., Zhu and Beroza, 2019). We found that this threshold offers a good balance between minimizing false picks and retaining valid arrivals, especially in low-SNR conditions. We added this explanation and a reference in the revised text (lines 99-102).

L83–L84: The choice of a minimum of 7 phases is not clearly explained. Can you give a bit more background or reasoning for this?

The reviewer is right, thanks, some elaboration is needed. After some tests, we chose a minimum of 7 picks as a good compromise between limiting false positives and detecting small earthquakes.

Furthermore, we apply a slightly more restrictive threshold compared to the CMS catalog, which uses a minimum number of 5. Since PhaseNet is more sensitive than the STA/LTA method and generally provides a higher number of picks for the same event, we opted for a higher threshold to ensure robustness and reduce the inclusion of low-quality detections. We now clarify this point in the manuscript (lines 108-111).

L84–L85: Is the 2 second residual cutoff in REAL a default value?, or did you choose it based on your results? A bit more explanation here would help readers follow your workflow.

The 2-second residual threshold is not the default setting in the REAL association algorithm (Zhang et al., 2019). We chose this relatively generous threshold to include more candidate events in the initial association step, which is useful when working with low-magnitude events and a sensitive phase picker like PhaseNet. A more selective filtering is then applied in the later stages of the workflow (the location with NonLinLoc, where events with RMS larger than 1 second are discarded — as described in the manuscript). Following the Reviewer’s suggestion, we clarified this point in the revised manuscript (lines 112-116).

L100–Eq. 1: I believe the parameters in this equation are standard for California, right? It probably works well enough, but can you add some support or justification for using it here as a proxy for ML?

We thank the Reviewer for this remark, and now motivate this choice in the manuscript. We use this equation because it is the same local magnitude (ML) relationship used for the reference CMS catalog for the Val d’Agri region. Although originally developed for earthquakes in California, this relationship has been chosen by the CMS team as the relation for computing local magnitude in this specific region. Therefore, we apply the same formula to ensure consistency and comparability with the reference catalog. We added a clarification and the appropriate citation (Morelli et al., 2025) in the revised manuscript (line 135).

L105: The CMS catalog also uses the same equation and parameters for ML? I think you mention this later, but it might help to say it earlier for clarity.

Indeed, the CMS catalog uses the same magnitude formula and parameters. As suggested, we moved this clarification earlier in the text to avoid confusion and improve clarity (line 134-136).

L133–L150: You justify the filtering using semblance–magnitude and average station distance, which is good. But I’d suggest adding a few waveform examples in the Supplementary, showing low-magnitude events that passed the filters. That would help readers evaluate the quality.

Done. We added some representative waveforms in the Supplementary information (Figure S4).

L161: The phrase “To compare catalog, we compare...” should be rephrased to improve the reading flow.

We thank the Reviewer for the suggestion again. We rephrased the sentence in the new version (line 224-225):

To assess the consistency between the catalogs, we analyze the time and locations of detected earthquakes, identifying matching patterns of epicenter clustering.

L167–L168: What do you mean by “high quality from NonLinLoc”? Can you add some quantitative explanation of what makes these locations better?

Thank you for your comment. We have revised the manuscript to clarify what we mean by “high quality from NonLinLoc.” While the absolute location uncertainties from NonLinLoc are generally larger than those provided by simpler methods, they are more realistic and statistically meaningful. This is because NonLinLoc employs a probabilistic approach based on grid search and travel-time probability density functions, which better capture the complexity and non-linearity of the earthquake location problem — even when using a 1D velocity model.

Additionally, we emphasize that the PRN catalog (NonLinLoc-based) provides higher spatial resolution than the CMS catalog, which reports event coordinates with limited decimal precision. While this coarser precision is consistent with the expected localization error, it prevents the identification of fine-scale spatial patterns in seismicity that could reflect underlying geological structures.

To clarify this point, we have added the following sentence to the Discussion (lines 273-276):

Although the absolute uncertainties from NonLinLoc are slightly larger, they provide a more reliable estimate of location accuracy due to the probabilistic nature of the algorithm, even when using a 1D velocity model. In contrast, the CMS catalog truncates hypocentral coordinates to fewer decimal places, reducing spatial resolution and limiting the ability to resolve small-scale structures illuminated by seismicity.

L183–L185: This paragraph could be rephrased for clarity.

We rephrased the paragraph to improve clarity as follows (lines 251-254):

Figure 5 shows the timeline of detected events and their magnitudes across the three catalogs. Overall, the magnitude estimates and their temporal evolution are consistent among the catalogs. For larger events, the magnitude estimates remain in good agreement. The main differences appear at lower magnitudes, where both QS and PRN detect additional smaller events, effectively lowering the detection threshold.

L197: When you say “attention zone,” do you mean the whole study area? Maybe use clearer wording here.

Exactly. “Attention zone” is the term employed to indicate the target area, defined by regulations to be subject to monitoring. We modified the term to “study area” to avoid misleading terms.

L201: You mentioned uncertainties from NonLinLoc earlier. Could you discuss general location uncertainties a bit more? That would help understand the quality of the catalogs’ location.

Thank you for this helpful suggestion. We have now included a more detailed discussion of the location uncertainties associated with each catalog at lines 228-233 (Results) and 273-276 (Discussion).

L239: Typo: “antropogenic” should be “anthropogenic”.

Corrected, thanks!

L270–L274: This paragraph is a bit unclear. The phrase “slight increase in seismicity” followed by “number of earthquakes remains low” is confusing. Try rephrasing and tie it more directly to what’s shown in Figure 8.

Following the Reviewer’s suggestion, we modified the paragraph as follows to improve clarity (lines 348-351):

Notably, in 2023, we observed both the highest water levels and the highest seismicity rate. In contrast, during 2022, although the reservoir also reached high levels, the seismicity remained low and comparable to the background level, with only a slight and short-lived increase. Overall, the analysis of our more complete seismic catalog supports the hypothesized seasonal pattern of local seismicity (Valoroso et al., 2023).

Figures

Figure 2: The color code by depth works well in the plain map view, but not in the cross-sections (since depth is already on the y-axis). You could use a plain color or color by something else in the cross-section. Also, I think the circle size reflects magnitude, if so, a legend mentioning that is needed. A colorbar is also missing for the plain view. Maybe add the number of events shown too? Also, the x-axis label is missing in the cross-sections; something like “Distance along profile” would work.

We have modified the figure, following the suggestion by the Reviewer. While we acknowledge that using color to represent depth in the vertical cross-sections may seem redundant (as depth is

already shown on the y-axis), we decided to retain the color scale to provide a visual cue in place of a separate legend, thus avoiding overloading the figure with additional graphical elements. To improve the readability of lighter-colored symbols, especially at shallow depths, we added a black edge to all circles. We also included a legend to the plain view indicating that circle size reflects magnitude, labeled the x-axis in the cross-sections as “Distance along profile”.

Figure 3: Please add a colorbar and the same point mentioned for Figure 2.

We have modified the figure, following the suggestion by the Reviewer.

We express our sincere gratitude to the Editor and both Reviewers for their thoughtful and constructive feedback; their comments have been fully addressed, leading to what we believe is a significant improvement of the manuscript.

Round 2

Reviewer A:

I would like to thank the authors for addressing my comments and questions. I would recommend this manuscript for publication after addressing the following minor comments.

Lines 140-141: How do you deal with stations without P or without S picks? Do you use a predicted arrival or just use those stations with both P and S?

Line 157-160: When there are gaps at some stations, maybe only a few stations (e.g. only 1 or 2 stations) have recordings. At the periods with gaps, maybe the semblance can readily exceed the threshold, resulting in locations with large uncertainties. How do you deal with this situation?

Lines 183-187: Thank you for manually checking the precision of Qseek-based workflow. This information is very useful. Have you considered doing the same thing to assess the precision of the PRN workflow? Although manual inspection is painful and dull, it might look weird if you only estimate the precision of the Qseek-based workflow in the paper.

Line 118, Line 367 and Data and code availability: Indent the first line.

Reviewer B:

Dear Editor,

I have reviewed the authors' responses and the revised version of the manuscript entitled "Improving detection of micro-earthquakes in the Val d'Agri region (Southern Italy) using deep learning algorithms." I appreciate the efforts they have made to address the comments and improve the manuscript. The additional references and expanded explanations contribute to a clearer and more complete presentation of the work.

The study represents a valuable contribution to the integration of deep learning methods in seismic monitoring workflows, with relevant implications for both research and operational applications. I am satisfied with the revisions and have no further comments. I recommend the manuscript for publication in its current form.

Response to reviewers

Improving detection of micro-earthquakes in the Val d'Agri region (Southern Italy) using deep learning algorithms

Caredda E., Isken M.P., Cesca S., Errico M., Zerbinato G., Morelli A.

September 8, 2025

Dear Editor and Reviewers,

We would like to thank you for your careful reading of our manuscript and for the valuable comments provided. We have carefully addressed all points raised in the second round of review. In the following, we provide a detailed point-by-point response to each comment, highlighting the changes made in the revised manuscript. We believe that these revisions have improved the clarity, completeness, and robustness of our study.

1 Reviewer A

I would like to thank the authors for addressing my comments and questions. I would recommend this manuscript for publication after addressing the following minor comments.

Lines 140-141: How do you deal with stations without P or without S picks? Do you use a predicted arrival or just use those stations with both P and S?

We thank the Reviewer for giving us the opportunity to explain this detail. The following considerations have been added to the revised manuscript. In our PRN workflow, stations without any pick, as well as stations with only S picks, are not considered. We only use stations with a P pick (alone or together with an S pick). If only the P pick is available, the amplitude window is anchored to the observed P arrival, while the S arrival is estimated approximately using constant P and S velocities ($V_p=6$ km/s and $V_s=3.5$ km/s following Dziewonski, 1981):

$$t_s \approx t_p + \Delta \left(\frac{1}{V_s} - \frac{1}{V_p} \right) \quad (1)$$

where Δ is the hypocentral distance. We then pad the window from -0.5 s before the P wave arrival to t_s+3s to ensure the S-wave coda is included, and compute the ML amplitude on the horizontal Wood–Anderson traces. If both P and S picks are available, the observed S is used; otherwise the predicted S is used. Stations without any pick are not used.

Line 157-160: When there are gaps at some stations, maybe only a few stations (e.g. only 1 or 2 stations) have recordings. At the periods with gaps, maybe the semblance can readily exceed the threshold, resulting in locations with large uncertainties. How do you deal with this situation?

In our implementation of *Qseek*, event detection requires a minimum of three stations. Windows with fewer stations are discarded, preventing spurious high semblance values from locations with very few recordings. This detail has been explicitly clarified in the revised manuscript.

Lines 183-187: Thank you for manually checking the precision of Qseek-based workflow. This information is very useful. Have you considered doing the same thing to assess the precision of the PRN workflow? Although manual inspection is painful and dull, it might look weird if you only estimate the precision of the Qseek-based workflow in the paper.

Thank you for the suggestion. We have now extended the manual precision check also to the PRN workflow. For the PRN workflow, the precision was $\sim 52\%$ in the initial output and increased to $\sim 88\%$ after applying the filtering steps. For the QS workflow, the corresponding values are $\sim 55\%$ (initial) and $\sim 78\%$ (final). We have added these results to the revised manuscript. They show that both workflows achieve a substantial improvement in precision after filtering, leading to a robust final catalog.

Line 118, Line 367 and Data and code availability: Indent the first line.

Done, thanks.

We thank Reviewer A once again for the constructive feedback and for the helpful suggestions that have improved the clarity and robustness of our manuscript.

2 Reviewer B

Dear Editor,

I have reviewed the authors' responses and the revised version of the manuscript entitled "Improving detection of micro-earthquakes in the Val d'Agri region (Southern Italy) using deep learning algorithms." I appreciate the efforts they have made to address the comments and improve the manuscript. The additional references and expanded explanations contribute to a clearer and more

complete presentation of the work.

The study represents a valuable contribution to the integration of deep learning methods in seismic monitoring workflows, with relevant implications for both research and operational applications.

I am satisfied with the revisions and have no further comments. I recommend the manuscript for publication in its current form.

Recommendation: Accept Submission

We thank Reviewer B for the positive evaluation and for their careful reading of our manuscript.

We appreciate their constructive comments and are glad that the revisions have improved the clarity and completeness of our work.

Round 3

Reviewer A:

The authors have addressed my comments.