Round 1

Reviewer A:

The paper describes QuakeMigrate, a flexible software for detecting and locating earthquakes directly from waveforms, and showcases its performance with synthetic data and two use cases at different scales. Whereas the QuakeMigrate software has been distributed as open source software for a few years and has been used successfully by many practioners, such a more formal description in the form of a Seismica manuscript is nevertheless highly appreciated.

The manuscript is very well executed, with a very clear description of the algorithm and the performance evaluation, excellent writing and informative figures. I was particularly impressed with the github repository allowing easy reproduction of all the results presented in this paper. I only tried the first figure but this went smoothly, including the installation of QuakeMigrate.

I only have very minor comments detailed below, with the main point being that the introductory review of similar approaches should include some earlier foundational work (see below or details).

Minor comments

p 1-2 Introduction: Include reference to the original source scanning algorithm publications.

The otherwise nicely written introduction relies entirely on publications published during the last 10-15 years, but the idea of migrating seismic energy is in fact older than that. A classic paper is the source-scanning paper of Kao and Shan:

Kao, H. & Shan, S.-J. (2004) The source-scanning algorithm: mapping the distribution of seismic sources in time and space Geophys. J. Int., , 157, 589-594

Some other older papers on this and related approaches are:

Baker, T.; Granat, R. & Clayton, R. W. (2005) Real-time earthquake location using Kirchhoff reconstruction, *Bul. Seism. Soc. Am.,* 95, 699-707

Kao, H.; Shan, S.; Dragert, H.; Rogers, G.; Cassidy, J.; Wang, K.; James, T. & Ramachandran, K. (2006) Spatial-temporal patterns of seismic tremors in northern Cascadia *J. Geophys. Res. Solid Earth*, 111 https://doi.org/10.1029/2005JB003727

It seems fair to acknowledge at least Kao & Shan's contribution with a citation and also include "source-scanning" in the list of synonyms in 63-65 on p 2.

P 2

L 37 "and the advent of storing continuous waveforms." – this has been the practice for 30+ years, so I am not sure it serves as a specific motivator for the developments. Consider deleting this half-sentence.

P 5

l 125 This may … → These may ...

p 6

l 162-163 You can refer here to the Acknowledgements where many contributors are listed, or simply state how many contributors QuakeMigrate had had.

P 7-8, Figure 5

In section 3.4 you state that you fit a 3D Gaussian function but in Fig. 5, it looks like the Gaussian ellipse is aligned with the coordinate axis, whereas the depth section in panel b seems to show the probability density function pattern to be tilted. It would be preferable to fit a 3D tilted ellipsoid, accounting for the covariances between latitude, longitude and depth errors.

P 7

l 203/204 “When combined using an appropriate stacking function, the resulting 3D-coalescnece maps can be interpreted as […] probability density maps.“ - this statement is unnecessarily vague. Effectively they can only be interpreted as (pseudo)probabilities if the individual (pseudo-probability) station map values are multiplied with each other, equivalent to stacking the log values. You describe this later in the manuscript in more detail, but still the statement here could be clearer.

P 8

Figure 3c: The caption states that the trace shown is the root-mean-square (rms) of the north and east components, but the rms should be positive everywhere while the panel shows positive and negative excursions. Please double-check.

P 9

l 233. The justification for doing calculations by summing logs rather than multiplying straight values is described as computational efficiency but the concern is really the precision of the calculation rather than efficiency. I would rephease the relative clause as “.., which, though eq 2 is mathematically equivalent, is computationally more robust through avoiding overflow.

P 10

l 285 “intermittent telemetry dropouts” – should only be a concern for realtime data as data is generally stored locally on the datalogger also. The data should be backfilled, when the connection comes back up again, or the station storage is read out directly. In my experience, gaps are more likely to come from clock problems or intermittent power outages.

P 14

l 362-364 Can you expand on your preference for geometric mean. A more physically intuitive measure would have been the root of the trace of the covariance matrix, i.e., \sqrt(\sigma_x^2 + \sigma_y^2 + \sigma_z^2)

l 375 Regarding your statement on false negative rates. I only understood from reading the supplement, how you define the ground truth for these claims, but ideally the main text should be understandable without referring to the supplement. Maybe you can mention briefly that you define the detection ground truth by comparing against a small subset that was manually inspected.

p 15

l 7 Somewhat off-topic for the paper, but I wonder how the streaks perpendicular to the flow direction arise.

p 19

l 465 Typo in QuakeMigrate

p 21

l 485 Not that it matters greatly, if the speedup is 100x, 50x, or 20x (not to forget that the analyst is free to work on something else, while the computer is working on detecting and locating earthquakes) but for a network of 20 stations, a 20 minutes per event time would imply picking takes nearly a minute per station. This seems excessively long, when an efficient manual picking program is used.

P 23 Data and code availability

It is appreciated that you note the DOIs of the data sets used, but the FDSN recommendation is to cite them equivalent to papers, i.e. you can cite “White (2007)” in the text for the 4F network and then put

Robert White. (2007). Askja 2007 [Data set]. International Federation of Digital Seismograph Networks. https://doi.org/10.7914/SN/4F_2007

in the References.

For Z7 the citation is

Robert White. (2010). Northern Volcanic Zone [Data set]. International Federation of Digital Seismograph Networks. https://doi.org/10.7914/SN/Z7_2010

You also mention the YG network, which presumably does not have a DOI. As YG is a temporary network code, it will be associated with different, usually unrelated experiments. To unambiguously describe the dataset, please add the year range or at least the starting year.

Supplementary material:

Section S1

p 1 bottom: I don't expect it makes a great deal of difference and I am not asking you to redo the calculation, but maybe you can comment why you assumed a constant angle of incidence rather than computing it (it could be computed from the gradient of the travel time grid near the station, although the surface response would also need to be taken into account.

Section S2 and Figure S3: add the fact that this is a synthetic example to the caption. Also, you can show the true location and origin time on the maps, and Coalescence-vs-time plots, as well as mention and discuss differences between true and inferred location in the text. While the uncertainties are reduced by the additional information in the S waves, the effect in the specific example seems rather marginal, especially for depth. I would imagine that the effect of adding S waves will be much more dramatic if you chose an event a little outside your array, and with the smallest epicentral distance larger than the hypocentral depth.

Also, it could be interesting to see for comparison a data example for a tectonic earthquake from Iceland with maybe messier arrivals.

Section S4,

p 6 2nd paragraph: The AUC definition for the precision-recall curve is not as commonly used as for ROC curves, and I have seen online resources that define the AUC as actual area under the curve even for precision-recall curves. I think you use the definition:

$$AUC = (AUC0 - PF) / (1 - PF)$$

with AUC0 the actual area under the Precision-recall curve and PF is the positive fraction, i.e. 117/139 ~ 84%. With this definition indeed an unskilled classifier would have an AUC value of 0 and a perfect one an AUC value of 1, but I think negative AUC could theoretically be possible. Please expand your description to provide the definition of the AUC in your case. Precision and Recall are commonly used metrics when the number of positive classes is much smaller than the number of negative examples. However, in your case it is the other way round and I wonder if a standard ROC curve and corresponding AUC value might be more intuitive but leave this for you to judge.

Figure S9: You mention that the "green dashed line shows the precision of the starting catalogue." I think this phrasing is not very clear. I suggest something like "green dashed line shows the proportion of positive samples (real events) in the sample, corresponding to the expected precision-recall curve for a completely unskilled (random) estimator."

Reviewer E:

Dear Authors,

Please find my review report attached.

Regards

General Remarks

The manuscript describes an open-source tool and its workflow for scanning continuous seismic waveforms, constructing characteristic functions (CFs) for individual stations, applying partial waveform stacking (PWS) for back-projection, generating a preliminary event catalog on a coarse spatial grid, and subsequently performing higher-resolution stacking to refine candidate detections. While the workflow is clearly presented, several concerns arise regarding the contribution and novelty of the work. The methodology primarily integrates existing and somewhat traditional techniques (e.g., STA/LTA-based CFs), yet the manuscript does not sufficiently justify the specific tool choices nor demonstrate clear advantages over comparable approaches. Similar frameworks have been developed in recent years, which incorporate advanced deep learning phase pickers and waveform-based detection strategies with demonstrated performance in dense waveform scanning and catalog compilation. A more direct comparison to these tools is needed to establish the novelty and performance gains of QuakeMigrate.

Furthermore, a more in-depth evaluation would strengthen the manuscript, including quantitative comparisons (e.g., detection rates, completeness curves, false positive rates) and at least a semi-quantitative assessment of uncertainties associated with the migration results. Without such analysis, it is difficult to assess whether the proposed workflow provides a meaningful improvement over existing approaches. I present my comments below:

Major comments:

1- The use of STA/LTA-based characteristic functions is a notable limitation of the current workflow. STA/LTA is generally less effective for detecting low-magnitude events and for data recorded in noisy environments. In addition, STA/LTA tends to perform efficient for identifying P-wave onsets than S-wave arrivals. However, accurate event location and reliable catalog construction require consistent and precise picking of both P and S phases. Given the recent development and widespread adoption of advanced deep learning–based phase detection methods (e.g., PhaseNet, EQTransformer, GPD), which have shown substantial improvements in sensitivity and picking accuracy, it is unclear why such approaches were not considered or incorporated into the workflow.

2- Does QuakeMigrate apply partial waveform stacking (PWS) to all continuous waveforms without any preliminary event detection using low-cost methods? Efficient phase association approaches, such as PyOcto applied to detected phases, can generate a preliminary event list at very low computational cost, substantially reducing the overall processing time. It is unclear why the current workflow imposes such high computational demands on the entire waveform dataset.

3- There have been several recently developed integrated workflows for seismic event detection and location, such as LOC-FLOW (Zhang et al., 2022, SRL), QuakeFlow (Zhu et al., 2023, GJI), IPIML (Mohammadigheymasi, 2023), and Malmi & Shi (2019). In particular, IPIML (Mohammadigheymasi, 2023) and Malmi & Shi (2019) employ a similar approach, incorporating advanced deep learning phase pickers and waveform-based detection strategies, and demonstrate strong performance in dense waveform scanning and catalog compilation. A more thorough discussion and quantitative comparison with these existing workflows would strengthen the manuscript.

4- Event locations should include uncertainty estimates. For the pick-based locations, hypoinverse and Nonlinloc provides uncertainties. For the migration-based locations, uncertainty could be quantified using a bootstrap approach or by analyzing the spatial extent of the focusing volume (e.g., using a fixed percentage of the maximum CF amplitude or an absolute threshold, given that the CFs are normalized). I recommend applying a bootstrap uncertainty analysis.

5- Some discussion of the *b*-value and magnitude of completeness for the different catalogs would be helpful in further evaluating the performance of the proposed workflow.

6- Figure 1: A clearer correspondence between the stations shown in the left panel and the P- and S-arrival time series in the middle panel is needed to more effectively illustrate the workflow.

7- Figure 2 is not sufficiently representative in its current form. The diagram should be revised to more clearly convey the workflow logic and the relationships between the processing steps.

Response to reviewers' comments on "QuakeMigrate: a Python Package for Automatic Earthquake Detection and Location Using Waveform Migration and Stacking" – Seismica manuscript #1854 by Winder, Bacon et al.

We thank the two reviewers for their constructive comments, which we have carefully considered in the revised version of the manuscript. Responses (in blue text) to each comment (in black text) are provided below, noting that line numbers refer to the marked-up version of the manuscript.

Reviewer A:

The paper describes QuakeMigrate, a flexible software for detecting and locating earthquakes directly from waveforms, and showcases its performance with synthetic data and two use cases at different scales. Whereas the QuakeMigrate software has been distributed as open source software for a few years and has been used successfully by many practioners, such a more formal description in the form of a Seismica manuscript is nevertheless highly appreciated.

The manuscript is very well executed, with a very clear description of the algorithm and the performance evaluation, excellent writing and informative figures. I was particularly impressed with the github repository allowing easy reproduction of all the results presented in this paper. I only tried the first figure but this went smoothly, including the installation of QuakeMigrate.

I only have very minor comments detailed below, with the main point being that the introductory review of similar approaches should include some earlier foundational work (see below or details).

We thank the reviewer for their careful review of the text and appreciation for the objective of this publication. We are particularly grateful for the suggestions for wording improvements to aid clarity, and ideas for further improvement to the software.

Minor comments

- p 1-2 Introduction: Include reference to the original source scanning algorithm publications.

  The otherwise nicely written introduction relies entirely on publications published during the last 10-15 years, but the idea of migrating seismic energy is in fact older than that. A classic paper is the source-scanning paper of Kao and Shan:
  - Kao, H. & Shan, S.-J. (2004) The source-scanning algorithm: mapping the distribution of seismic sources in time and space Geophys. J. Int., 157, 589-594

  Some other older papers on this and related approaches are:
  - Baker, T.; Granat, R. & Clayton, R. W. (2005) Real-time earthquake location using Kirchhoff reconstruction, Bul. Seism. Soc. Am., 95, 699-707

- Kao, H.; Shan, S.; Dragert, H.; Rogers, G.; Cassidy, J.; Wang, K.; James, T. & Ramachandran, K. (2006) Spatial-temporal patterns of seismic tremors in northern Cascadia J. Geophys. Res. Solid Earth, 111 https://doi.org/10.1029/2005JB003727

It seems fair to acknowledge at least Kao & Shan's contribution with a citation and also include "source-scanning" in the list of synonyms in 63-65 on p 2.

Thank you for pointing out this omission. We aimed to provide a short summary and draw on the more detailed account of Li et al. (2020), but you are correct that these fundamental earlier papers should be included, as well as explicit reference to "source-scanning".

➔ Added at Lines 59-73. Apologies that the added references are not accurately rendered as tracked changes.

- P 2 L 37 "and the advent of storing continuous waveforms." – this has been the practice for 30+ years, so I am not sure it serves as a specific motivator for the developments. Consider deleting this half-sentence.

We agree with the reviewer's point that referring to the "advent" of storing continuous waveform data is an over-emphasis and have changed the text appropriately. In our experience this has not been widespread practice for quite such a long period - in the case of the second example dataset used here, several stations from the Icelandic Meteorological Office are not included in light of them still sending triggered event data until 2012.

➔ We have changed: "advent of storing…" → "widespread adoption of storing continuous waveforms instead of triggered event data" (Lines 37-38).

- P 5 L 125 This may … → These may ...

➔ Corrected. (Line 126).

- p 6 L 162-163 You can refer here to the Acknowledgements where many contributors are listed, or simply state how many contributors QuakeMigrate had had.

➔ Thank you for this suggestion – reference to the Acknowledgments has been added. (Line 166).

- P 7-8, Figure 5

In section 3.4 you state that you fit a 3D Gaussian function but in Fig. 5, it looks like the Gaussian ellipse is aligned with the coordinate axis, whereas the depth section in panel b seems to show the probability density function pattern to be tilted. It would be preferable to fit a 3D tilted ellipsoid, accounting for the covariances between latitude, longitude and depth errors.

The reviewer is correct that the uncertainty ellipse shown in Fig. 5 and, more generally, throughout the package, is aligned with the coordinate axes. At present, the software outputs only the axis-aligned marginal uncertainties in X, Y, and Z of a Cartesian space, which is consistent with the conventions used in other established tools and maintains backwards compatibility with existing workflows.

Internally, however, the fitting procedure does estimate a fully rotated 3-D Gaussian, including the covariances between all spatial dimensions. We agree that exposing and visualising the full ellipsoid—its principal axes, orientation, and principal standard deviations—would provide a more complete representation of the probability density functions. Such a change would require some revision of the structure of the outputs from the Locate stage of QuakeMigrate, which requires some careful consideration to minimise backwards-compatibility issues, and a minor release. We have marked this as an important enhancement and hope to include it in a new version of the package in 2026.

- P 7 L 203/204 “When combined using an appropriate stacking function, the resulting 3D-coalescnece maps can be interpreted as […] probability density maps.“ - this statement is unnecessarily vague. Effectively they can only be interpreted as (pseudo)probabilities if the individual (pseudo-probability) station map values are multiplied with each other, equivalent to stacking the log values. You describe this later in the manuscript in more detail, but still the statement here could be clearer.

Thank you for this suggestion. Text updated:

➔ “When combined using an appropriate stacking function, the resulting..” → “When combined through multiplication, the resulting 3-D coalescence maps can be interpreted as (non-normalised) probability density functions describing the source location uncertainty”. (Lines 206-207).

- P 8 Figure 3c: The caption states that the trace shown is the root-mean-square (rms) of the north and east components, but the rms should be positive everywhere while the panel shows positive and negative excursions. Please double-check.

This is because only one of the horizontal components is plotted in panel c. This has now been made clear in the figure caption:

➔ Added “Note that for clarity only one of the two horizontal components is shown in (c).” (Figure 3 caption).

- P 9 L 233. The justification for doing calculations by summing logs rather than multiplying straight values is described as computational efficiency but the concern is really the precision of the calculation rather than efficiency. I would rephrase the relative clause as “.., which, though eq 2 is mathematically equivalent, is computationally more robust through avoiding overflow.

➔ Changed as suggested. (Lines 234-238).

- P 10 L 285 “intermittent telemetry dropouts” – should only be a concern for realtime data as data is generally stored locally on the datalogger also. The data should be backfilled, when the connection comes back up again, or the station storage is read out directly. In my experience, gaps are more likely to come from clock problems or intermittent power outages.

Though this is true for modern, sophisticated dataloggers, the authors’ experience is that there are present-day networks (e.g. at AVO in Alaska and IMO in Iceland) either still heavily leveraging dataloggers without this capability, or not prioritising (or even enabling) direct download from the datalogger – including within the dataset used for the Iceland example. The reviewer is correct that power outages and clock problems are other common causes of intermittent dropouts, but – where they still occur – telemetry dropouts are significantly more impactful as a large proportion of stations are more likely to be affected.

- P 14 L 362-364 Can you expand on your preference for geometric mean. A more physically intuitive measure would have been the root of the trace of the covariance matrix, i.e., \sqrt(\sigma_x^2 + \sigma_y^2 + \sigma_z^2)

Thank you for this suggestion. We appreciate the reviewer’s point that this measure is more familiar, and perhaps also more stable in the case of very low covariance values along at least one axis. We have tested this option for the icequake example presented in the manuscript (please see Figures A, B & C below). In practice this shows that the performance is extremely similar to the geometric mean, which is perhaps unsurprising, given their similar characteristics. We take this suggestion as a valuable alternative to incorporate in future, but given the similar outcome of these two measures we do not make any change to what is presented here.

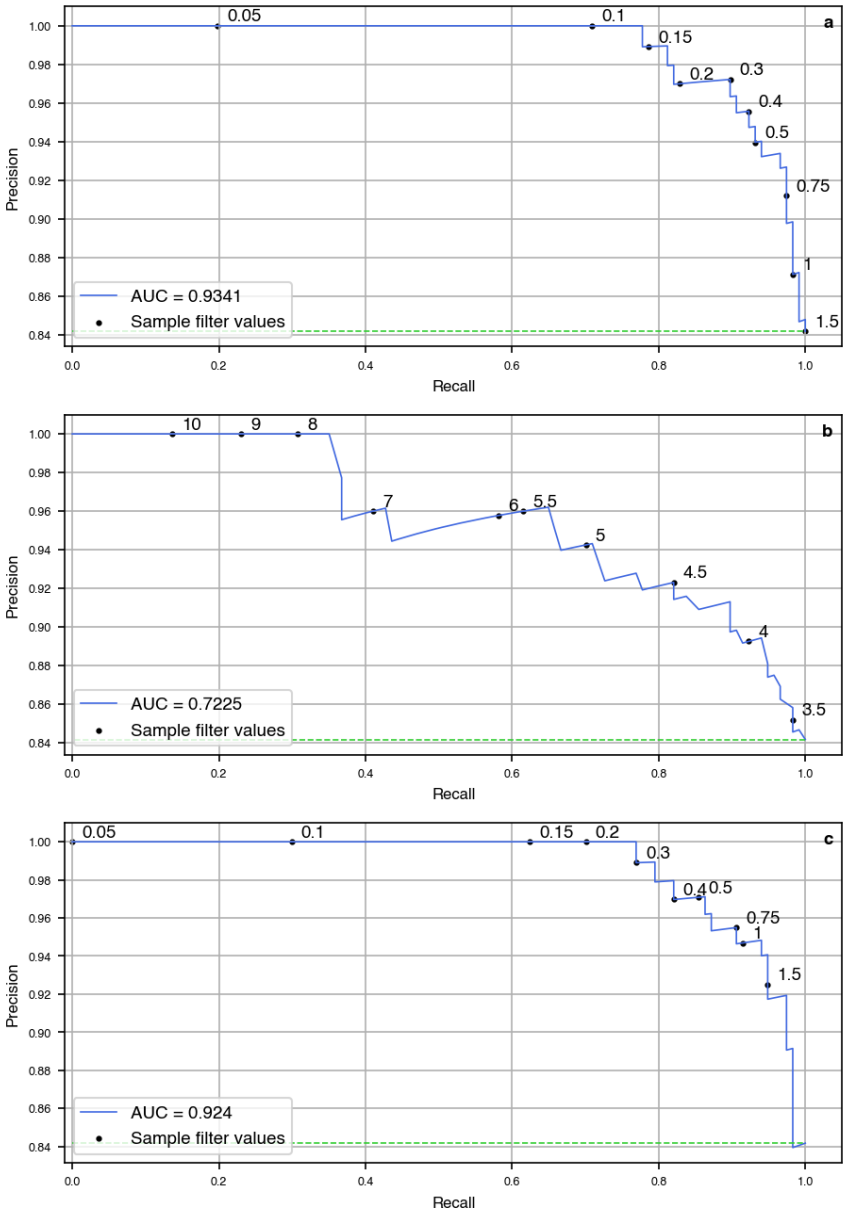

Figure A: As Figure S9 (now S11; see caption there for reference), but with panel **c** showing Global Covariance calculated as the root of the trace of the covariance matrix. Performance is similar to **a**.

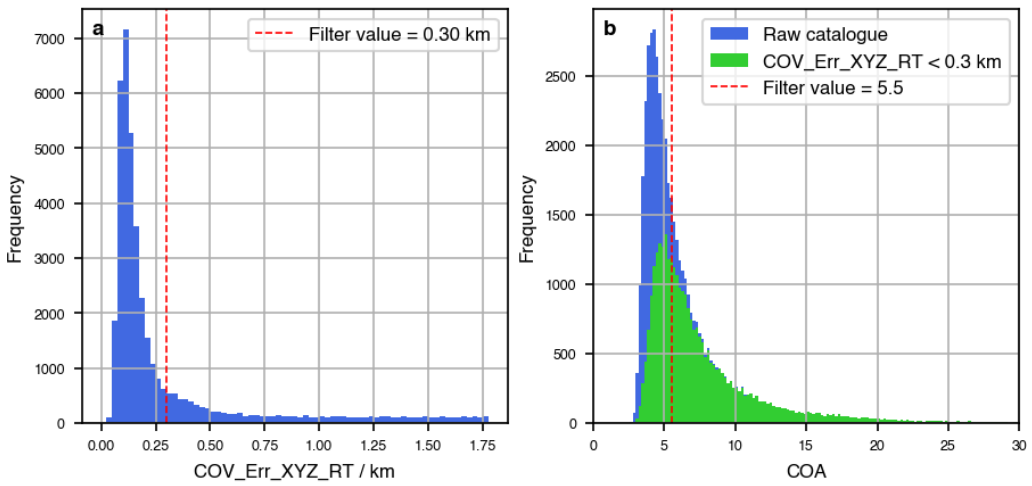

Figure B: As Supplementary Figure S10 (now S12; please see caption for reference), but using the root of the trace of the covariance matrix to calculate the Covariance statistic. A comparable filter value has been chosen (0.3 km) according to the labelled Precision-Recall curve in Figure A.

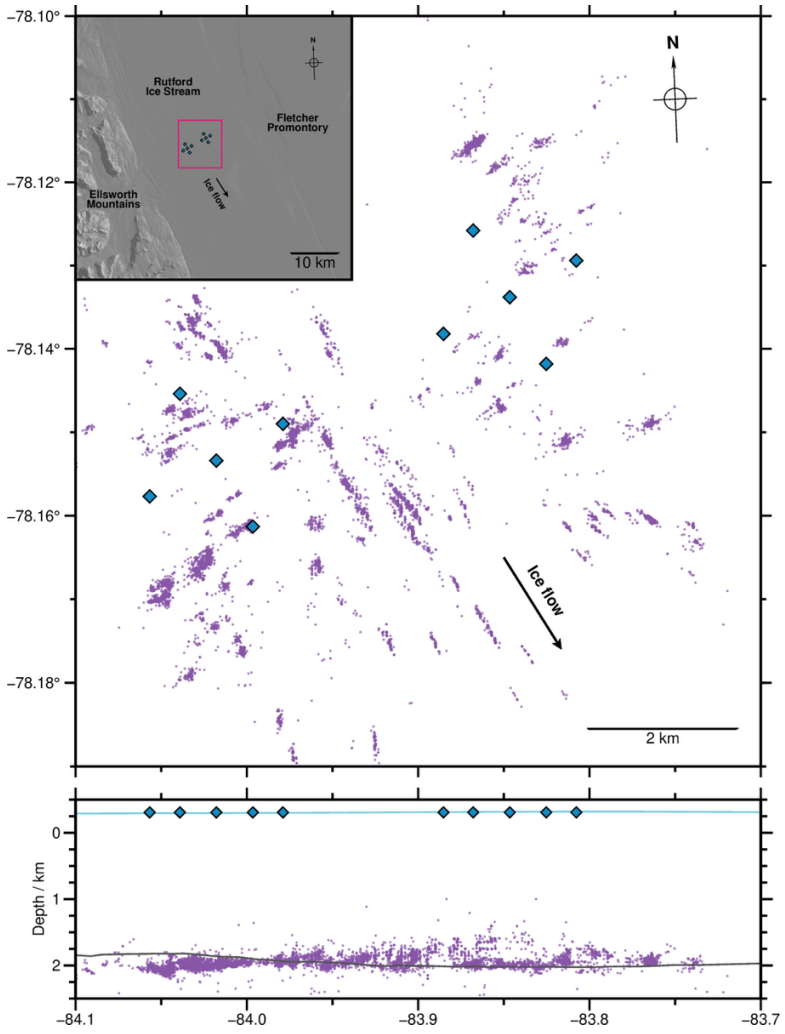

Figure C: As Figure 6, but using a filter based on the root of the trace of the covariance matrix (0.3 km, and COA >= 5.5). 20,000 events are retained (in comparison to 20,015 originally), and qualitatively the geometry of structures etc. is unchanged.

- P14 L 375 Regarding your statement on false negative rates. I only understood from reading the supplement, how you define the ground truth for these claims, but ideally the main text should be understandable without referring to the supplement. Maybe you can mention briefly that you define the detection ground truth by comparing against a small subset that was manually inspected.

Thank you for this suggestion – we have modified the text to clarify this:

➔ Added "In order to make a more informed choice of filter value, we perform a statistical analysis of a representative subset of manually labelled events - comprising the first 10.5 minutes of the dataset, containing 149 / 47,354 events – categorised based on inspection of the Locate summary plots. Based on this analysis, we choose to remove events with …". Lines 368-373.

➔ Added "…for these filter values, the comparison to the ground-truth subset of manually labelled events indicates that the expected proportion of true positives …". Lines 380-381.

- p 15 L 7 Somewhat off-topic for the paper, but I wonder how the streaks perpendicular to the flow direction arise.

Good observation. There could be several causes of this. A glaciological reason could be that icequakes nucleate along paths of weakness. Typically these are oriented in the ice flow direction, but we have new evidence (work currently in review) that sometimes the hydrological pressure gradient can drive seismicity parallel to topographic gradient, which in some areas is perpendicular to ice flow. In some cases, there is only a slight streaking that looks radial relative to the centroid of a single array. In these cases, it is likely due to location uncertainty caused by the imperfect array geometry meaning that locations are dominated by a single back-azimuth (e.g. akin to beamforming errors).

- p 19 L 465 Typo in QuakeMigrate

➔ Corrected. (Line 472).

- p 21 L 485 Not that it matters greatly, if the speedup is 100x, 50x, or 20x (not to forget that the analyst is free to work on something else, while the computer is working on detecting and locating earthquakes) but for a network of 20 stations, a 20 minutes per event time would imply picking takes nearly a minute per station. This seems excessively long, when an efficient manual picking program is used.

This was estimated from personal experience – though admittedly generally on networks of 40-100 stations – and factoring in reviewing a wadati plot etc. to quality-check the picks, in addition to inspecting residuals from the location inversion, and revising or removing errant picks, if necessary. As prefaced by the reviewer, we do not think it is important (nor possible) to give a truly general estimate here, so we prefer to retain the current estimate, which in our experience is representative of the dataset in question.

- P 23 Data and code availability

  It is appreciated that you note the DOIs of the data sets used, but the FDSN recommendation is to cite them equivalent to papers, i.e. you can cite “White (2007)” in the text for the 4F network and then put

  Robert White. (2007). Askja 2007 [Data set]. International Federation of Digital Seismograph Networks. https://doi.org/10.7914/SN/4F_2007

  in the References.

➔ Reference updated. (Line 575).

- For Z7 the citation is:

  Robert White. (2010). Northern Volcanic Zone [Data set]. International Federation of Digital Seismograph Networks. https://doi.org/10.7914/SN/Z7_2010

➔ Reference updated. (Line 575).

- You also mention the YG network, which presumably does not have a DOI. As YG is a temporary network code, it will be associated with different, usually unrelated experiments. To unambiguously describe the dataset, please add the year range or at least the starting year.

Thank you for pointing this out.

➔ Reference to “YG” has been corrected to “YG_2009”. (Line 574).

Supplementary material:

- Section S1, P 1 bottom: I don't expect it makes a great deal of difference and I am not asking you to redo the calculation, but maybe you can comment why you assumed a constant angle of incidence rather than computing it (it could be computed from the gradient of the travel time grid near the station, although the surface response would also need to be taken into account.

The constant angle of incidence was initially used as a simple way of mixing some of the P- and S-phase arrival energy across the 3 components of the synthetic sensors. In response to this comment, we have adjusted the synthetic waveform generation script to use more realistic incidence angles for the given event-network geometry. To do so, we used NonLinLoc to calculate take-off angles at each station, which are determined from gradients of travel-time at the node, as suggested by the reviewer.

➔ An additional figure showing the variation in incidence angle as a function of radial distance from the earthquake source has been added to the supplementary material (Figure S2). This necessarily changes all subsequent figures that are derived from the synthetic waveforms (Figures 1, 3, 4, 5 & S4), but they do not materially change any of the results presented in the manuscript.

- Section S2 and Figure S3: add the fact that this is a synthetic example to the caption. Also, you can show the true location and origin time on the maps, and Coalescence-vs-time plots, as well as mention and discuss differences between true and inferred location in the text. While the uncertainties are reduced by the additional information in the S waves, the effect in the specific example seems rather marginal, especially for depth. I would imagine that the effect of adding S waves will be much more dramatic if you chose an event a little outside your array, and with the smallest epicentral distance larger than the hypocentral depth.

  Also, it could be interesting to see for comparison a data example for a tectonic earthquake from Iceland with maybe messier arrivals.

For the synthetic example we do not show the location and origin times on the map & cross-sections and coalescence vs time plot, because they are near indistinguishable, as noted by the reviewer. However, we have added an additional figure showing a similar comparison for one event from the Askja example, which shows a much clearer location difference. Here we do annotate the "P-only" figure with the reference location and origin time from the P- and S-phase derived location, as suggested.

➔ Added reference to the Figure S3 (now S4) caption that this is a synthetic example.
➔ Added reference to the location differences to the text of section S2.
➔ Added additional Figure S5, showing a similar comparison for an event from the Askja example, annotated with reference location & origin time
➔ Added discussion of the Askja event comparison to the text of section S2. Added text is: "… for the synthetic example, in Figure S4, and for one event from the Askja example, in Figure

S5. For the synthetic example the location difference is only 0.05 km laterally and 0.2 km in depth, and the origin time is identical, which highlights that for very well constrained event locations (with clear phase arrivals and excellent network geometry) P-phase only locations can perform well, though the location uncertainty is still larger. In contrast, excluding the S-phase onset functions makes a substantial difference for the event from the Askja example (Figure S5), where the location is offset by 7.5 km laterally, and 0.3 km in depth, as well as a 0.72 s difference in origin time. The estimated location uncertainty is also significantly larger in all three dimensions - and particularly in depth - as well as the geometric mean covariance, corresponding to the significantly worse location constraint. This example makes it clear that for smaller events, and where network geometry is not optimal, using all available phases is essential to obtain a reliable location.

- Section S4, P 6 2nd paragraph: The AUC definition for the precision-recall curve is not as commonly used as for ROC curves, and I have seen online resources that define the AUC as actual area under the curve even for precision-recall curves. I think you use the definition:

  AUC = (AUC0 – PF) / (1 – PF)

  with AUC0 the actual area under the Precision-recall curve and PF is the positive fraction, i.e. 117/139 ~ 84%. With this definition indeed an unskilled classifier would have an AUC value of 0 and a perfect one an AUC value of 1, but I think negative AUC could theoretically be possible. Please expand your description to provide the definition of the AUC in your case. Precision and Recall are commonly used metrics when the number of positive classes is much smaller than the number of negative examples. However, in your case it is the other way round and I wonder if a standard ROC curve and corresponding AUC value might be more intuitive but leave this for you to judge.

The reviewer is correct in their understanding of the modified AUC calculation we use. Thank you for highlighting this ambiguity, which we have now clarified, by adding:

➔ “For each case an AUC (area under the curve) statistic is calculated as

$$AUC = \frac{AUC_0 - PREC_i}{1 - PREC_i}$$

where AUC_0 is the full trapezoidal area under the Precision-Recall curve, and $PREC_i$ is the precision of the input catalogue (the proportion of “real” events) corresponding to the expected score for a completely unskilled (random) classifier). The AUC score therefore indicates the filtering performance above random chance.” Supplementary Section S4.

Regarding the choice of Precision-Recall (PR) vs Receiver Operating Curve (ROC) analysis, we share the understanding that the primary benefit of using PR comes where the sample is strongly skewed, and particularly where there is a very small proportion of positive samples. However, we also see the benefit that the concepts of Precision and Recall translate more directly to the balance of priorities for a user in this case typically seeking to find a compromise between catalogue “clean-

ness" vs. completeness. We believe the modified AUC definition still provides an intuitive measure of classification performance (comparable to the ROC AUC score), notwithstanding the possibility of obtaining a negative value. We therefore choose not to change this.

- Figure S9: You mention that the "green dashed line shows the precision of the starting catalogue." I think this phrasing is not very clear. I suggest something like "green dashed line shows the proportion of positive samples (real events) in the sample, corresponding to the expected precision-recall curve for a completely unskilled (random) estimator."

Corrected as suggested:

➔ Added: "i.e. the proportion of positive samples ("real" events) in the set of labelled events, corresponding to the expected Precision-Recall curve for a completely unskilled (random) classifier". Figure S9 (now S11) caption.

Reviewer E:

General Remarks

The manuscript describes an open-source tool and its workflow for scanning continuous seismic waveforms, constructing characteristic functions (CFs) for individual stations, applying partial waveform stacking (PWS) for back-projection, generating a preliminary event catalog on a coarse spatial grid, and subsequently performing higher-resolution stacking to refine candidate detections. While the workflow is clearly presented, several concerns arise regarding the contribution and novelty of the work. The methodology primarily integrates existing and somewhat traditional techniques (e.g., STA/LTA-based CFs), yet the manuscript does not sufficiently justify the specific tool choices nor demonstrate clear advantages over comparable approaches. Similar frameworks have been developed in recent years, which incorporate advanced deep learning phase pickers and waveform-based detection strategies with demonstrated performance in dense waveform scanning and catalog compilation. A more direct comparison to these tools is needed to establish the novelty and performance gains of QuakeMigrate.

Furthermore, a more in-depth evaluation would strengthen the manuscript, including quantitative comparisons (e.g., detection rates, completeness curves, false positive rates) and at least a semi-quantitative assessment of uncertainties associated with the migration results. Without such analysis, it is difficult to assess whether the proposed workflow provides a meaningful improvement over existing approaches. I present my comments below:

We thank the reviewer for reading the manuscript and providing their comments. However, we find some of their suggestions are to add items that are already included in the submitted manuscript. We regret that the reviewer missed or misunderstood these sections on first reading, and attempt to clarify these points in detail below. Regarding the contribution and novelty of this work, we refer to the Seismica guidelines for a "Software Report", which state:

- the primary goals are "*to document new codes, to facilitate community use of them, and to ensure reproducibility of their outputs*"
- the main paper "*should describe the scientific context and the methods employed, and should detail aspects such as test case simulations and model verification, evaluation, and performance*"

We believe we satisfy these requirements. Moreover, we firmly believe that open-source software tools deserve recognition for their usability and the facility they provide for open, reproducible research, in addition to providing advances in performance. This is described in detail in the paper, exhibited in the open and accessible nature of all code, dependencies and documentation, and further demonstrated by the wide variety of research which has already been carried out using QuakeMigrate, a sample of which is listed below:

1. Reiss, M.C., Muirhead, J.D., Laizer, A.S., Link, F., Kazimoto, E.O., Ebinger, C.J. and Rümpker, G., 2021. The impact of complex volcanic plumbing on the nature of seismicity in the developing magmatic natron rift, Tanzania. *Frontiers in Earth Science*, *8*, p.609805.
2. Hudson, T.S., Baird, A.F., Kendall, J.M., Kufner, S.K., Brisbourne, A.M., Smith, A.M., Butcher, A., Chalari, A. and Clarke, A., 2021. Distributed acoustic sensing (DAS) for natural microseismicity

studies: A case study from Antarctica. *Journal of Geophysical Research: Solid Earth*, *126*(7), p.e2020JB021493.

3. Widiyantoro, S., Supendi, P., Ardianto, A., Baskara, A.W., Bacon, C.A., Damanik, R., Rawlinson, N., Gunawan, E., Sahara, D.P., Zulfakriza, Z. and Husni, Y.M., 2022. Implications for fault locking south of Jakarta from an investigation of seismic activity along the Baribis fault, northwestern Java, Indonesia. *Scientific reports*, *12*(1), p.10143.
4. Glastonbury-Southern, E., Winder, T., White, R.S. and Brandsdóttir, B., 2022. Ring fault slip reversal at Bárðarbunga volcano, Iceland: Seismicity during caldera collapse and re-inflation 2014–2018. *Geophysical Research Letters*, *49*(21), p.e2021GL097613.
5. Greenfield, T., Winder, T., Rawlinson, N., Maclennan, J., White, R.S., Ágústsdóttir, T., Bacon, C.A., Brandsdóttir, B., Eibl, E.P., Glastonbury-Southern, E. and Gudnason, E.Á., 2022. Deep long period seismicity preceding and during the 2021 Fagradalsfjall eruption, Iceland. *Bulletin of Volcanology*, *84*(12), p.101.
6. Hudson, T.S., Brisbourne, A.M., Kufner, S.K., Kendall, J. and Smith, A.M., 2023. Array processing in cryoseismology: a comparison to network-based approaches at an Antarctic ice stream. *The Cryosphere*, *17*(11), pp.4979-4993.
7. Dando, B.D., Goertz-Allmann, B.P., Brissaud, Q., Köhler, A., Schweitzer, J., Kværna, T. and Liashchuk, A., 2023. Identifying attacks in the Russia–Ukraine conflict using seismic array data. *Nature*, *621*(7980), pp.767-772.
8. Hudson, T.S., Kufner, S.K., Brisbourne, A.M., Kendall, J.M., Smith, A.M., Alley, R.B., Arthern, R.J. and Murray, T., 2023. Highly variable friction and slip observed at Antarctic ice stream bed. *Nature Geoscience*, *16*(7), pp.612-618.
9. Supendi, P., Winder, T., Rawlinson, N., Bacon, C.A., Palgunadi, K.H., Simanjuntak, A., Kurniawan, A., Widiyantoro, S., Nugraha, A.D., Shiddiqi, H.A. and Adi, S.P., 2023. A conjugate fault revealed by the destructive Mw 5.6 (November 21, 2022) Cianjur earthquake, West Java, Indonesia. *Journal of Asian Earth Sciences*, *257*, p.105830.
10. Muñoz-Burbano, F., Calò, M., Savard, G., Reyes-Orozco, V. and Lupi, M., 2024. Using time-lapse seismic velocity changes to monitor the Domo de San Pedro Geothermal field, Mexico. *Geothermics*, *120*, p.103010.
11. Lee, I.R., Anandakrishnan, S., Alley, R.B., Brisbourne, A. and Smith, A., 2024. Characterization of Sticky Spots in Rutford Ice Stream, West Antarctica with High-Granularity Microseismicity. *Authorea Preprints*.
12. Agnew, R.S., Pearce, E., Karplus, M., Ranganathan, M., Hoffman, A.O., Hunt, M., Pretorius, A., Shanly, S.E., Beres, M., Pradhan, K.K. and Seldon, Y., 2025. Active and Passive Seismic Surveys over the Grounding Zone of Eastwind Glacier, Antarctica. *Seismological Research Letters*.
13. Paap, B., Vandeweijer, V., van Wees, J.D. and Kraaijpoel, D., 2025. Leveraging Distributed Acoustic Sensing for monitoring vessels using submarine fiber-optic cables. *Applied Ocean Research*, *154*, p.104422.
14. Yemane, T., Hudson, T.S., Kendall, J.M., Blundy, J., Tadesse, A.Z., Hammond, J.O., Ayele, A., Ogubazghi, G. and Lapins, S., 2025. Interconnectivity of magmatic and hydrothermal systems of Aluto volcano in the Main Ethiopian Rift inferred from seismicity. *Journal of Geophysical Research: Solid Earth*, *130*(6), p.e2024JB031053.
15. Glastonbury-Southern, E., Winder, T., Rawlinson, N., White, R.S., Greenfield, T., Bacon, C.A., Ágústsdóttir, T., Brandsdóttir, B., Gudnason, E.Á., Hersir, G.P. and Fischer, T.J., 2025. Pre-existing structures control the orientation of strike-slip faulting during the 2021 dike intrusion at Fagradalsfjall, Iceland. *Journal of Geophysical Research: Solid Earth*, *130*(6), p.e2024JB030162.

Major comments:

1. The use of STA/LTA-based characteristic functions is a notable limitation of the current workflow. STA/LTA is generally less effective for detecting low-magnitude events and for data recorded in noisy environments. In addition, STA/LTA tends to perform efficient for identifying P-wave onsets than S-wave arrivals. However, accurate event location and reliable catalog construction require consistent and precise picking of both P and S phases. Given the recent development and widespread adoption of advanced deep learning–based phase detection methods (e.g., PhaseNet, EQTransformer, GPD), which have shown substantial improvements in sensitivity and picking accuracy, it is unclear why such approaches were not considered or incorporated into the workflow.

We draw the reviewer's attention to section 3.4 "Calculation of onset functions". Here we justify the use of the STA/LTA characteristic function as the basis (default) option – with references including Beskardes et al. (2018), who provide an objective comparison of a variety of options, within a framework directly applicable to our use-case. The reviewer states that "*STA/LTA is generally less effective for detecting low-magnitude events and for data recorded in noisy environments*", however they do not provide any reference to support this claim. Through the two example use-cases included in the manuscript, we directly address this issue by demonstrating the performance of a well-parameterised STA/LTA onset function in exactly these environments. In all examples we use STA/LTA to successfully identify both P- and S-phase arrivals (e.g. see Figure 3). We have also added (also in response to a comment from Reviewer #1) a new supplementary figure (S5), and expanded the discussion in Supplementary material section S2, which more clearly demonstrates the advantage of using both P- and S-phase onset functions, as we successfully achieve with the default STA/LTA onset function.

Notwithstanding the choice of *default* characteristic function, we thank the reviewer for highlighting the importance of providing a piece of software where direct comparisons can readily be made between different characteristic functions, to expand on the analysis of Beskardes et al.. We direct them to Lines 211-222, from which we quote "*Advances in phase identification are continuously being made, with ever more information from the raw signal (e.g., phase, frequency, and polarisation) being incorporated, both analytically (references omitted) and via machine learning approaches (references omitted). In anticipation of continued future improvements, this module has been implemented as a 'plug-in' such that any such transformation might be used, and rigorously compared within a single framework"* and Lines 531-536 of the Conclusions. We hope it is clear upon re-reading these sections that such approaches as the reviewer refers to have been considered (all suggested references are already included in the above quoted section from Lines 211-218), and that these are readily incorporated into the workflow.

2. Does QuakeMigrate apply partial waveform stacking (PWS) to all continuous waveforms without any preliminary event detection using low-cost methods? Efficient phase association approaches, such as PyOcto applied to detected phases, can generate a preliminary event list at very low computational cost, substantially reducing the overall processing time. It is

unclear why the current workflow imposes such high computational demands on the entire waveform dataset.

As detailed throughout sections 1 and 2, there are clear performance advantages (in terms of detection capability) that accompany the moderately increased computational cost of the PWS approach. One novelty of the implementation provided within QuakeMigrate is to reduce the processing time through dividing the workflow into 3 stages, where the goals of detection are disentangled from those of location – see Lines 137-145. The suggestion to use an alternative method for "*preliminary event detection*" – particularly through using PyOcto, which operates on picks – implies a regrettable misunderstanding of the foundational motivation for the PWS workflow: to avoid the necessity to make picks on records from individual seismic stations as the first step in event detection, thereby losing the advantages offered by exploiting the coherence of information from many closely-spaced stations within a seismic network. These advantages are demonstrated in the successful detection (& location) performance exhibited upon application to the two example datasets – we draw particular attention to Figures 6, 7, 8 & 9, and the text of Sections 4.1 and 4.2.

If a user wishes to only use the "Locate" module of QuakeMigrate, then they may, if they wish, provide a list of events (detected with PyOcto or any other software of their choice) in substitute of the "Trigger" output – this is intentionally made possible through the modular nature of the workflow, and the transparent inputs and outputs to each run stage – and is utilised in Section 4.2 (Lines 469-472) to construct the location benchmark (Figure 10). We prefer to look at this in a positive light: QuakeMigrate offers a valuable new option for a user to apply to their dataset (whether they use the full pipeline or only certain segments). It will not provide the best compromise for all users or all datasets, but in this paper we aim to outline the motivation for its development, and describe potential advantages which we hope many users will benefit from.

3. There have been several recently developed integrated workflows for seismic event detection and location, such as LOC-FLOW (Zhang et al., 2022, SRL), QuakeFlow (Zhu et al., 2023, GJI), IPIML (Mohammadigheymasi, 2023), and Malmi & Shi (2019). In particular, IPIML (Mohammadigheymasi, 2023) and Malmi & Shi (2019) employ a similar approach, incorporating advanced deep learning phase pickers and waveform-based detection strategies, and demonstrate strong performance in dense waveform scanning and catalog compilation. A more thorough discussion and quantitative comparison with these existing workflows would strengthen the manuscript.

We agree with the reviewer's sentiment that a comprehensive comparison of the wide variety of automated and semi-automated workflows for earthquake detection and location would be a valuable resource to establish, however we consider this beyond the scope of the present manuscript.

To address the comment more specifically, LOC-FLOW and QuakeFlow are implementations of purely "pick then locate" approaches, implementing phase pickers, associators, and location algorithms which we already refer to in the submitted manuscript. MALMI (also referred to

specifically in the text) leverages the LOKI migration software (Grigoli et al., 2013, 2016, 2018 – referenced throughout the manuscript) for event location, but not detection. The IPIML workflow uses a pick-then-locate workflow, leveraging MALMI for the location step. Though these workflows all have their own distinguishing features, we believe it is beyond the scope of this manuscript introducing and outlining the QuakeMigrate software to go into the minutiae of the numerous previously published workflows (of which there are many more than suggested here). We refer to our response to reviewer 2's point #1, where we highlight the sections of the manuscript where we discuss the potential benefits in using onset functions derived from machine-learning, as implemented by MALMI, where Shi et al. (2022) is appropriately referenced.

Furthermore, we note that in the four references the reviewer provides (each outlining a new piece of software, or analysis pipeline), none of them include a quantitative comparison against multiple pre-existing options, as is being requested here. Instead, they each present a real-world usage example where they refer to a reference catalogue for assessment of detection and/or location performance. Here we present two real-world usage examples, including one with a reference catalogue (Section 4.1, Smith et al. (2015) Lines 380-385, Figures 6 & 7; Section 4.2, Lines 439-489, Figures 8, 9 & 10). Accordingly, we do not make any change to the manuscript here, though we look forward to a future publication focussed on benchmarking across detection & location workflows.

4. Event locations should include uncertainty estimates. For the pick-based locations, hypoinverse and Nonlinloc provides uncertainties. For the migration-based locations, uncertainty could be quantified using a bootstrap approach or by analyzing the spatial extent of the focusing volume (e.g., using a fixed percentage of the maximum CF amplitude or an absolute threshold, given that the CFs are normalized). I recommend applying a bootstrap uncertainty analysis.

The event locations produced by QuakeMigrate do include uncertainty estimates. This is discussed extensively in the manuscript, and plotted in Figures 5, 9, S4, S5, and S13. See Lines 16 (in the abstract), 142, 204-210, 223-226, 310-315, 434-437. Our preference for estimating the location uncertainty directly, as opposed to using post-hoc statistical treatments such as bootstrapping, is specifically addressed in Lines 208-210. A second form of spatial uncertainty estimate (the "global covariance" metric) is calculated by taking the covariance of the spatial volume clipped above a certain percentage of the normalised coalescence peak, as described in the text in Lines 362-368, corresponding to the second suggestion made here by the reviewer to analyse the spatial extent of the focussing volume.

5. Some discussion of the b-value and magnitude of completeness for the different catalogs would be helpful in further evaluating the performance of the proposed workflow.

It is possible to calculate a b-value and magnitude of completeness for the Askja example (section 4.2), but for a catalogue of 39 events this is not meaningful. We do not calculate earthquake

magnitudes for the Rutford catalogue (this is possible but not attempted here – see e.g. Kufner et al. 2021). However, in any case this only provides an insight into relative detection performance against a reference catalogue; the absolute value of Mc is dominantly controlled by noise conditions, event characteristics and network geometry. We therefore do not undertake this.

6. Figure 1: A clearer correspondence between the stations shown in the left panel and the P- and S-arrival time series in the middle panel is needed to more effectively illustrate the workflow.

The waveforms in the middle panels of Figure 1, and in all panels of Figure 3, have been reordered by distance. The distance is measured from the proposed source to the receiver, the effect of which is to make the ordering of traces between the upper and lower central panels of Figure 1 different (distance in the upper panel is measured from the top-left of the search grid, whilst in the lower panel it is measured from the centre of the search grid, i.e., the true location). A note has been added to the figure caption to explain.

- ➔ Revised station ordering in Figure 1, and updated Figure 1 caption to make this clearer.
- ➔ Figure 3 ordering is also revised to maintain correspondence with the bottom row of Figure 1 (which shows the correct source location).

7. Figure 2 is not sufficiently representative in its current form. The diagram should be revised to more clearly convey the workflow logic and the relationships between the processing steps.

We regret that the reviewer does not find the current version of this figure sufficiently representative. However, without raising any specific points, it is difficult to know what exactly they find unclear. We have considered several options to revise the figure, but find the current layout best suited to conveying the package structure (which is its primary aim). Adding further annotation to also detail the workflow logic confuses the layout, which we believe does not provide a net benefit. We instead refer the reviewer to the documentation (https://quakemigrate.readthedocs.io) for details on the workflow logic, where we provide several worked examples as well as accompanying commentary on parameter choices and trade-offs, and the relationships between stages of a QuakeMigrate run. It is beyond the scope of this paper outlining the method to provide such details, which we believe are better provided in the “living” documentation, such that they can be kept updated as the package evolves.

Round 2

Reviewer A:

For author and editor

The authors answered nearly all the points raised by me to my full satisfaction. I still think that some of the network practices described by them as the standard ~15 years ago, i.e., use of triggered data and data gaps due to telemetry dropouts, were no doubt still present in a number of networks and seem to have affected some of the networks they were working with, but even at that time that was not state-of-the-art. But this is a very marginal point, and I will not insist on any further change. See below for the one remaining very minor suggestion I have.

In contrast, the comments of reviewer 2 were addressed w\ith a rebuttal arguments but usually to refuse the suggestions for being out-of-scope. I agree with the authors' assessment, however, that reviewer 2 made unreasonable demands given the purpose of the paper is to introduce a particular software, which is behind many scientific publications already, rather than a comprehensive review and benchmarking of migration-based algorithms.

My last remaining actionable comment regards Figure 3 caption:

>> P 8 Figure 3c: The caption states that the trace shown is the root-mean-square (rms) of the north and east components, but the rms should be positive everywhere while the panel shows positive and negative excursions. Please double-check.

> This is because only one of the horizontal components is plotted in panel c. This has now been

made clear in the figure caption:

> → Added "Note that for clarity only one of the two horizontal components is shown in (c)."

(Figure 3 caption).

Thank you for adding this clarification, which helps, but you still have the phrase " (depicted as the root-mean-square of the north and east components)", which to me means that you actually show this. A minor rephrase similar to the following would be sufficient to remove ambiguity: "S-wave onset functions (panel d) are derived in the same manner from the representative horizontal component trace, calculated as the root-mean-square of the north and east components) (panel c)."

Reviewer B:

For author and editor

Dear Authors

I thank you for the efforts in addressing the comments and provide the following feedback on the responses; while some comments were addressed convincingly, I offer additional feedback on a few points. The limitations of the STA/LTA method for phase detection, particularly its reduced effectiveness for low magnitude events, data recorded in noisy environments, and its comparatively weaker performance in identifying S-wave arrivals relative to P-wave onsets, are well documented in the literature (e.g., [1], [2]). The authors address this issue by suggesting that appropriate hyperparameter selection can improve performance; however, hyperparameter tuning itself introduces an additional challenge, as implicitly demonstrated by the two comprehensive examples presented in the manuscript. This point is especially relevant given that the current software implementation relies exclusively on STA/LTA as the phase-picking engine, despite acknowledging the potential inclusion of alternative methods. Accordingly, I note that the manuscript would benefit from a dedicated **Limitations** section explicitly discussing this issue, which would further strengthen the manuscript.

Despite the detailed explanation provided in the manuscript regarding the division of the workflow into three stages to reduce computational time, a quantitative comparison of computational cost is still missing. In particular, a computational cost curve comparing the waveform processing time required by a conventional deep learning–based phase picker (e.g., EQTransformer or PhaseNet) with that of QuakeMigrate for the **detection step only** would be valuable. I consider this to be an additional limitation of the current workflow that should be explicitly acknowledged and discussed in a dedicated **Limitations** section.

The update to Figure 1 has improved its readability and interpretability; however, Figure 2 still lacks sufficient clarity. The main issue with Figure 2 is that it remains overly abstract relative to the implemented workflow. In particular, the “three-stage detect and locate” concept is not clearly reflected in the implementation flow and would benefit from clearer integration with the actual workflow steps. As figures should be self-explanatory, I suggest expanding the figure caption to provide more detailed context and explicitly defining the LUT acronym within the caption to improve interpretability.

Minor comment:
In Equation (3), the comma should link the equation to the following text; starting a new paragraph after the equation is incorrect.

[1] Zhang, Y., Chen, Q., Liu, X., Zhao, J., Xu, Q., Yang, Y., & Liu, G. (2018). Adaptive and automatic P-and S-phase pickers based on frequency spectrum variation of sliding time windows. *Geophysical Journal International*, *215*(3), 2172-2182.
[2] Ho, L. M., Walter, J. I., Hansen, S. E., Sánchez-Roldán, J. L., & Peng, Z. (2024). Evaluating automated seismic event detection approaches: An application to Victoria Land, East Antarctica. *Journal of Geophysical Research: Machine Learning and Computation*, *1*(3), e2024JH000185.

Response to reviewers' comments on "QuakeMigrate: a Python Package for Automatic Earthquake Detection and Location Using Waveform Migration and Stacking" – Seismica manuscript #1854 by Winder, Bacon et al.

We thank the two reviewers for their constructive comments, which we have carefully considered in the revised version of the manuscript. Responses (in blue text) to each comment (in black text) are provided below, noting that line numbers refer to the marked-up version of the manuscript.

Reviewer A:

The authors answered nearly all the points raised by me to my full satisfaction. I still think that some of the network practices described by them as the standard ~15 years ago, i.e., use of triggered data and data gaps due to telemetry dropouts, were no doubt still present in a number of networks and seem to have affected some of the networks they were working with, but even at that time that was not state-of-the-art. But this is a very marginal point, and I will not insist on any further change. See below for the one remaining very minor suggestion I have.

In contrast, the comments of reviewer 2 were addressed with a rebuttal arguments but usually to refuse the suggestions for being out-of-scope. I agree with the authors' assessment, however, that reviewer 2 made unreasonable demands given the purpose of the paper is to introduce a particular software, which is behind many scientific publications already, rather than a comprehensive review and benchmarking of migration-based algorithms.

We thank the reviewer for reading our response to their comments and revised manuscript.

My last remaining actionable comment regards Figure 3 caption:

- P 8 Figure 3c: The caption states that the trace shown is the root-mean-square (rms) of the north and east components, but the rms should be positive everywhere while the panel shows positive and negative excursions. Please double-check.

  This is because only one of the horizontal components is plotted in panel c. This has now been made clear in the figure caption:

  > *→ Added "Note that for clarity only one of the two horizontal components is shown in (c)." (Figure 3 caption).*

  Thank you for adding this clarification, which helps, but you still have the phrase " (depicted as the root-mean-square of the north and east components)", which to me means that you actually show this. A minor rephrase similar to the following would be sufficient to remove ambiguity: "S-wave onset functions (panel d) are derived in the same manner from the representative horizontal component trace, calculated as the root-mean-square of the north and east components) (panel c)."

Thank you for this suggestion – we have improved the wording. The comment highlights that our explanation of the order of operations here was unclear: where multiple components are assigned to a single phase in the channel mapping (here both north and east for the S-phase), individual onset functions are first calculated for each component, before the root-mean-square of these per-

component onset functions are taken to give a single onset function for that phase. The same is true if e.g. a hydrophone channel is additionally specified for P-phase onset functions when dealing with OBS data.

➔ Changed “..are calculated in the same manner from the horizontal components (depicted as the root-mean-square of the north and east components) (panel c).” → “..are calculated in the same manner from each of the horizontal components (panel c) with a single combined S-wave onset function calculated as the root-mean-square of the north and east component onset functions, where the default channel mapping is used.” (Figure 3 caption)

Reviewer B:

Dear Authors

I thank you for the efforts in addressing the comments and provide the following feedback on the responses; while some comments were addressed convincingly, I offer additional feedback on a few points.

We thank the reviewer for reading the text again and providing their feedback.

The limitations of the STA/LTA method for phase detection, particularly its reduced effectiveness for low magnitude events, data recorded in noisy environments, and its comparatively weaker performance in identifying S-wave arrivals relative to P-wave onsets, are well documented in the literature (e.g., [1], [2]). The authors address this issue by suggesting that appropriate hyperparameter selection can improve performance; however, hyperparameter tuning itself introduces an additional challenge, as implicitly demonstrated by the two comprehensive examples presented in the manuscript. This point is especially relevant given that the current software implementation relies exclusively on STA/LTA as the phase-picking engine, despite acknowledging the potential inclusion of alternative methods. Accordingly, I note that the manuscript would benefit from a dedicated **Limitations** section explicitly discussing this issue, which would further strengthen the manuscript.

To address and the following point, we have introduced an explicit "Discussion" section in the manuscript, comprising the previous subsection 4.3 (now divided between 5.3 and 5.4), and two new subsections. The first new Discussion subsection ("5.1 – Alternative algorithms for onset function calculation") addresses the comparative strengths and weaknesses of the STA/LTA algorithm for onset function calculation compared to other statistical methods and particularly machine learning techniques.

- ➔ Added new subsection title 4.3 "Location benchmark: comparison with manually picked earthquakes at Askja" – previously this was included within subsection 4.2 (Line 456)
- ➔ Added new section title 5 "Discussion" (Line 494)
- ➔ Moved previous subsection 4.3 "Ongoing and future developments in automated workflows and seismic data acquisition" → 5.3 "Incorporation into automated workflows" and 5.4 "Ongoing and future developments in seismic data acquisition" (Lines 613 and 631)
- ➔ Added new subsection 5.1 "Alternative algorithms for onset function calculation" (Line 495)

  "The default onset function algorithm included in the QuakeMigrate package is based on the short-term average to long-term average ratio (STA/LTA) of the seismogram amplitude. The core migration and stacking routine and the onset function module are implemented with a well-defined interface, however, enabling alternative onset functions to be incorporated as plug-ins and directly compared. The STA/LTA algorithm is lightweight and requires minimal pre-processing, provides good resolution and performance even for small earthquakes (e.g., Beskardes et al., 2018; Cesca and Grigoli, 2015), and, importantly, can be applied to any seismic dataset without the requirement for, e.g., pre-training a suitable machine-learning model, or assembly of template events for matched-filter processing. Some degree of hyperparameter tuning is required to optimise performance for a given signal frequency content (Sections 4.1 and 4.2), however the individual parameters are intuitively linked to the signal properties (Hudson et al., 2019). Furthermore, it has been shown that the STA/LTA parameters used in Section 4.2 perform well across a broad range of tectonic and volcanic microseismicity, covering magnitudes ~ -2–5 (e.g., Greenfield et al., 2022; Supendi et al.,

2023; Hudson et al., 2022; Yemane et al., 2025)—including locating both the DLP and shallow tectonic earthquakes present in the Askja dataset shown here, which have significantly different frequency content (Figure 8)."

"Recently, significant progress has been made in developing a large family of phase detection algorithms based on machine-learning techniques (e.g., Ross et al., 2018; Mousavi et al., 2020; Zhu and Beroza, 2019). Unlike the STA/LTA algorithm, these can be designed to exploit information beyond the sharp increases in signal amplitude associated with primary seismic phase arrivals. This can help overcome the reduced sensitivity of STA/LTA where arrivals are emergent, for S-arrivals masked by the coda from earlier arriving P-waves, or for closely spaced events where the SNR is again reduced by the preceding coda (Schaff and Beroza, 2004). Furthermore, many of these techniques analyse 3-component data simultaneously, instead of using the simple mapping of vertical-component data to the P-wave onset function and horizontals to S described here (Section 3.4). This reduces the likelihood of errant peaks in the S-wave onset function caused by the presence of P-wave energy on the horizontal components (due to non-vertical incidence and/or scattering), and vice versa (e.g., Figure 3b), helping to avoid artefacts caused by this mistaken phase identification—though this can also be mitigated by utilising polarisation information} (e.g., Grigoli et al., 2014). It has also been shown that, where suitably trained, machine-learning phase detection algorithms can provide significantly improved robustness to data errors, including spikes caused by broken equipment, data gaps, and cultural noise which has significantly different time-frequency characteristics to earthquake phase arrivals, and avoid the tendency of STA/LTA algorithms to systematically pick slightly early (e.g., Mousavi et al., 2020)."

"Direct performance comparisons between STA/LTA and machine learning methods focus on their application for phase picking, instead of onset function calculation for migration. This is a fundamentally different application, where observations are only counted if they exceed a specified "pick threshold", whereas within the migration framework any peak that stacks coherently with other phase arrivals detected across the network will successfully contribute to the final detection/location (see Grigoli et al., 2018, for a more complete discussion). However, the general insight applies that—where appropriate hyperparameter choices are made, including the pick threshold—STA/LTA typically provides high recall (identifying most true earthquake phase arrivals) at the cost of lower precision when compared to the best-performing machine learning tools—i.e., exhibiting more false positives (e.g., Mousavi et al., 2020). However, this comparative disadvantage is strongly mitigated by the robust and implicit phase association capability of the migration and stacking approach implemented within QuakeMigrate (Sections 2 & 3.5). Furthermore, the recall capability of machine-learning algorithms is dependent on both their architecture and the data used for their training. Though built to generalise, they typically face challenges when applied to datasets that differ strongly from the catalogues on which they were trained (e.g., Lapins et al., 2021; Shi et al., 2024; Ho et al., 2024; Mousavi et al., 2020). This is particularly relevant in applications such as studying icequakes (Section 4.1), which represent an "out-of-distribution" generalisation, because most models are trained on datasets comprised dominantly of tectonic earthquakes (e.g., Ross et al., 2018; Zhu and Beroza, 2018; Mousavi et al., 2020; Shi et al., 2024) which have distinct characteristics. There are ongoing developments to address this, but so far these come at significantly increased computational cost (e.g., Shi et al., 2024), or require transfer learning on a labelled sample dataset (e.g., Lapins et al., 2021; Ho et al., 2024). Neither of these developments fully addresses the fact that, particularly in volcanic settings, new phases of unrest can produce previously unseen seismic signals that differ from existing earthquake templates (Lapins et al., 2021), which may therefore be missed by machine-learning or matched-filter algorithms, but still be successfully identified by simpler and less specific statistical algorithms such as STA/LTA, kurtosis, or the Hilbert transform (e.g. Shi et al., 2024).

It is likely that combining the complementary strengths of different onset function algorithms—or comparing their performance for a given dataset, to assess which best suits the user's needs—will provide the best outcome. By allowing alternative onset functions to be incorporated as plug-ins in place of the default STA/LTA algorithm, QuakeMigrate provides a common framework to achieve this with both presently available onset function algorithms and those yet to be developed."

(Lines 496-550)

We add the following references, including reference [2] mentioned by the reviewer:

1. Zhu W., & Beroza, G.C., (2018). PhaseNet: a deep-neural-network-based seismic arrival-time picking method. *Geophysical Journal International*, *216*(1) pp.261-273.
2. Schaff, D.P. and Beroza, G.C., 2004. Coseismic and postseismic velocity changes measured by repeating earthquakes. *Journal of Geophysical Research: Solid Earth, 109*(B10).
3. Shi, P., Meier, M.A., Villiger, L., Tuinstra, K., Selvadurai, P.A., Lanza, F., Yuan, S., Obermann, A., Mesimeri, M., Münchmeyer, J. and Bianchi, P., 2024. From labquakes to megathrusts: Scaling deep learning based pickers over 15 orders of magnitude. *Journal of Geophysical Research: Machine Learning and Computation*, *1*(4), p.e2024JH000220.
4. Ho, L.M., Walter, J.I., Hansen, S.E., Sánchez-Roldán, J.L. and Peng, Z., 2024. Evaluating automated seismic event detection approaches: An application to Victoria Land, East Antarctica. *Journal of Geophysical Research: Machine Learning and Computation*, *1*(3), p.e2024JH000185. .
5. Greenfield, T., Winder, T., Rawlinson, N., Maclennan, J., White, R.S., Ágústsdóttir, T., Bacon, C.A., Brandsdóttir, B., Eibl, E.P., Glastonbury-Southern, E. and Gudnason, E.Á., 2022. Deep long period seismicity preceding and during the 2021 Fagradalsfjall eruption, Iceland. *Bulletin of Volcanology*, *84*(12), p.101..
6. Supendi, P., Winder, T., Rawlinson, N., Bacon, C.A., Palgunadi, K.H., Simanjuntak, A., Kurniawan, A., Widiyantoro, S., Nugraha, A.D., Shiddiqi, H.A. and Adi, S.P., 2023. A conjugate fault revealed by the destructive Mw 5.6 (November 21, 2022) Cianjur earthquake, West Java, Indonesia. *Journal of Asian Earth Sciences*, *257*, p.105830..
7. Hudson, T.S., Kendall, J.M., Pritchard, M.E., Blundy, J.D. and Gottsmann, J.H., 2022. From slab to surface: Earthquake evidence for fluid migration at Uturuncu volcano, Bolivia. *Earth and Planetary Science Letters*, *577*, p.117268..
8. Yemane, T., Hudson, T.S., Kendall, J.M., Blundy, J., Tadesse, A.Z., Hammond, J.O., Ayele, A., Ogubazghi, G. and Lapins, S., 2025. Interconnectivity of magmatic and hydrothermal systems of Aluto volcano in the Main Ethiopian Rift inferred from seismicity. *Journal of Geophysical Research: Solid Earth*, *130*(6), p.e2024JB031053..

Of which #1-4 are used to expand the discussion of alternative onset functions, and their strengths and weaknesses, and #5-8 are used to support the statement that the STA/LTA onset function implemented in QuakeMigrate has been shown to be widely applicable with no or very modest hyperparameter tuning – in direct response to the reviewer's comment regarding the challenge this may present for a user.

Despite the detailed explanation provided in the manuscript regarding the division of the workflow into three stages to reduce computational time, a quantitative comparison of computational cost is still missing. In particular, a computational cost curve comparing the waveform processing time required by a conventional deep learning–based phase picker (e.g., EQTransformer or PhaseNet) with that of QuakeMigrate for the **detection step only** would be valuable. I consider this to be an

additional limitation of the current workflow that should be explicitly acknowledged and discussed in a dedicated **Limitations** section.

We have added a second new Discussion subsection ("5.2 – Computational cost") to provide a more detailed and quantitative account of the computational cost of the QuakeMigrate Detect stage, referring to both example use cases presented in Section 4. Note that this information was already included in the original submission, but only in the Supplementary information (Section S3) – here we simply add a discussion of it within the main text.

We remain of the opinion that to perform a detailed comparison – by applying a deep-learning based phase picking workflow to one of these datasets – is beyond the scope of this paper. To be meaningful, this comparison would require a quantitative comparison of the detection performance, as well as runtime, and given the wide variety of deep-learning pickers and phase association methods available (see e.g. Woollam et al., 2022 & Munchmayer et al., 2024) it is not representative to only choose one. Instead, we make a quantitative comparison using indicative runtimes for a range of machine learning pickers – drawing largely on the benchmarking work of Münchmeyer et al. (2022) – as well as for associators (again drawing on an indicative set of comparisons from Münchmeyer, 2024). This allows us to place the runtime of the Detect stage of QuakeMigrate into perspective of the runtime using a deep-learning workflow, as requested by the reviewer, which we agree is a useful information to provide in the manuscript.

➔ Added new subsection 5.2 "Computational cost" (Line 551)

> "The exhaustive 4-D grid search executed in the Detect stage of a QuakeMigrate run is a continued focus for efficiency improvements, to enhance usability across a wide range of possible applications: from centimetre-scale acoustic-emission experiments to regional seismic arrays. A detailed overview of the computational cost for the Askja usage example (Section 4.2), and its scaling with parameter choice, is provided in the Supplementary Material (Section S3). Here we do not consider the Trigger stage (which generally runs in << 1 minute on a single thread for one day of data; Supplementary Figure S8) or the Locate stage, which is similar in runtime to alternatives such as NonLinLoc (Section 4.3), and significantly shorter than the Detect step other than where the event count is exceptionally high—and which is also required in a pick-then-locate workflow. For the Askja example, the optimal parameter choice for Detect results in a runtime of 38 seconds for 1 hour of data using 4 threads, or approximately 15 minutes for a full day of data, and ~ 92 hours (3.8 days) for one year. Computational cost is strongly dependent on the number of nodes in the search grid, which is controlled by the grid extent (in x, y, and z), and node spacing (Supplementary Figure S7), as well as the sampling rate. The Askja example provides a representative use-case for a local seismic network, covering 71 x 62 x 40 km. When decimated by a factor of 2 for Detect, this gives a node count of 36 x 32 x 21, or 24,192 grid nodes, with potential origin times sampled at 50 sps. Searching at the same sampling rate for a search grid spanning 150 x 150 x 40 km would result in a Detect runtime of ~ 1 hour for a full day of data."

> "Where denser node spacing is required to achieve the level of resolution appropriate for very high frequency signals and dense network spacing, such as in the application to basal icequakes at Rutford ice stream (Section 4.1), computational cost is significantly higher, even for a much smaller grid extent. Here the LUT spans 9.2 x 10.05 x 1.5 km, but with a node spacing of 100 m for Detect, giving 150,288 nodes. Scanning at 250 sps, this results in a compute time ~ 30 times higher. The memory requirement (which would otherwise scale similarly) is mitigated by completing the scan in shorter timesteps."

“The icequake example therefore runs approximately 3 x faster than real-time, while the Askja example runs ~ 100 x faster. These runtimes can be compared to the now widely used family of machine-learning pickers, which generally run in a pick, associate, then locate workflow. Here, where the user has access to a suitable graphics processing unit (GPU) and is using a pre-trained model, phase identification and picking is relatively fast: of order 1–5 seconds per station per 24h of 3-component data (excluding GPD, which runs much more slowly), with most towards the bottom of that range (Münchmeyer et al., 2022; Lapins et al., 2021), or ~ 2.5–12 hours for the 23 stations used in the Askja example, for 1 year of data. In this workflow, however, the association step must also be considered for a representative comparison to Detect, and—particularly where many events are detected—this can be a computationally intensive task. In the Askja region, 32 events/day were located between 2007–2020 using QuakeMigrate (Winder, 2022). This totals ~ 12,000 events per year. We refer to the runtime comparisons in (Münchmeyer, 2024)—which are indicative, rather than a full-scale benchmark of seismic phase associators—to provide an indication of runtimes for this use-case. For their shallow seismicity synthetic benchmark, with 100 events per day (~ double that at Askja), runtimes range from 0.7–13.0 s / day, or 4–80 minutes for 1 year of data. For higher event rates, performance is significantly worse; the maximum event rate tested was 2,000 / day—significantly below the 10,000 events per day in even the filtered Rutford icequake catalogue (Section 4.2)—and in this case several associators did not complete the task within 48 hours for one day of data, though PyOcto and REAL ran within 0.3–20 minutes, depending on the number of mis-picks on noise (Münchmeyer, 2024). Indicative overall runtimes are therefore ~ 2.5–14 hours for the machine-learning pipelines, which is a significant factor faster than QuakeMigrate Detect, at ~ 92 hours.”

“It is important to put these runtimes in context of their typical use in a research workflow, where differences of hours to several tens of hours between catalogue generation jobs are generally not a barrier to adoption, and to acknowledge that it is common for users to have access to significantly more CPU resources than used here, which can be exploited through parallelisation for considerable speed-ups. Notably, the association benchmarks from Münchmeyer (2024) were run on 16 threads, while the QuakeMigrate runtimes are from processing on 4 threads only. Speed-ups for the migration do not continue to scale linearly above this due to I/O and other overheads (Supplementary Figure S7), but running multiple Detect jobs in parallel—e.g., by splitting a year of data into several batches—means one can continue to achieve near-linear speed-up, reducing the runtimes by a further factor of 4, and reaching similar computational cost to the machine-learning pipeline. On a server or cluster with 100 CPUs or more, QuakeMigrate would run significantly faster. This, too, is without requiring access to a machine with a dedicated GPU. In the current implementation, for dense earthquake sequences—at small spatial scale, and with short inter-event intervals—migration is an equivalent or more efficient option, however in sparse cases—with low spatial density of events, and/or long inter-event times—the pick-then-locate workflow strongly benefits from earlier data reduction, whereas retaining continuous onset functions and combining them in the migration search uses significant computational resources evaluating non-hypocentres. These two very different approaches are therefore complementary depending on the user's priorities and the application, and particularly in cases where a suitable pre-trained model is not available (Sections 4.1 & 5.1). Furthermore, there remain several opportunities to improve the efficiency of QuakeMigrate, including by sub-sampling the search grid with an Oct-Tree algorithm (Drew et al., 2013; Isken et al., 2025), implementing a boxcar filter to enable more aggressive decimation of the search grid without compromising detection performance (Drew et al., 2013), and decoupling the scan sampling rate from the onset function sampling rate, to search over possible origin times at a coarser interval (Shi et al., 2019). Together these can realistically achieve an order of magnitude decrease in runtime.”

(Lines 552-612)

We added the following reference, to provide a quantitative evaluation of a variety of deep-learning based phase pickers (as mentioned by the reviewer):

- Münchmeyer, J., Woollam, J., Rietbrock, A., Tilmann, F., Lange, D., Bornstein, T., Diehl, T., Giunchi, C., Haslinger, F., Jozinović, D. and Michelini, A., 2022. Which picker fits my data? A quantitative evaluation of deep learning based seismic pickers. *Journal of Geophysical Research: Solid Earth*, *127*(1), p.e2021JB023499.

The update to Figure 1 has improved its readability and interpretability; however, Figure 2 still lacks sufficient clarity. The main issue with Figure 2 is that it remains overly abstract relative to the implemented workflow. In particular, the "three-stage detect and locate" concept is not clearly reflected in the implementation flow and would benefit from clearer integration with the actual workflow steps. As figures should be self-explanatory, I suggest expanding the figure caption to provide more detailed context and explicitly defining the LUT acronym within the caption to improve interpretability.

Figure 2 is intended to show the package structure, instead of the workflow. To address this secondary need, we have added a new supplementary figure S1, which shows the step-by-step pipeline, including clearly depicting the three-stage detect/trigger/locate concept. This is referred to in the text (Lines 153, 186, 215, 254, 271, 292, 299, 390, 499). In order to further clarify Figure 2, we have made the following improvements in line with the reviewer's suggestions:

- ➔ We have added a definition of the LUT acronym to the caption of Figure 2, as suggested (Figure 2 caption). In addition, the definition is now included on the figure.
- ➔ Added new Supplementary Figure S1 (workflow schematic); note previous supplementary figure numbering has been updated, and also updated throughout the text.

Minor comment:
In Equation (3), the comma should link the equation to the following text; starting a new paragraph after the equation is incorrect.

Thank you for this suggestion. However, the formatting of Equation (3) is identical to the previous two equations, and we do not – at least as we read it – start a new paragraph afterwards. In any case, this is a typesetting issue, and we are simply following the journal template. If there is an issue, we trust it will be resolved in the proofs stage.

[1] Zhang, Y., Chen, Q., Liu, X., Zhao, J., Xu, Q., Yang, Y., & Liu, G. (2018). Adaptive and automatic P-and S-phase pickers based on frequency spectrum variation of sliding time windows. *Geophysical Journal International*, *215*(3), 2172-2182.
[2] Ho, L. M., Walter, J. I., Hansen, S. E., Sánchez-Roldán, J. L., & Peng, Z. (2024). Evaluating automated seismic event detection approaches: An application to Victoria Land, East Antarctica. *Journal of Geophysical Research: Machine Learning and Computation*, *1*(3), e2024JH000185.