**Reviewer A**

One general comment concerns the connection between the enhanced earthquake catalog and the DYFI analysis. I found the relationship between these two components somewhat difficult to follow. The enhanced catalog appears to be motivated by improving the characterization of the aftershock sequence and extracting additional information about the spatial and temporal evolution of seismicity. In contrast, the DYFI analysis is more directly related to population distribution and shaking intensity, as the authors themselves discuss. As currently presented, these two components feel somewhat disconnected. I wonder whether the manuscript presents only one component of a larger study, as several aspects of the sequence are introduced but not explored in detail. While I understand that additional analyses of this sequence may be ongoing

and reserved for future work, I found the overall scope of the manuscript somewhat fragmented.

We agree that the connection between the machine learning catalogue and the DYFI analysis could be clearer. Although the two datasets address different aspects of the earthquake sequence, the enhanced catalogue provides an important temporal and spatial framework for interpreting the macroseismic intensity observations. Having a more complete aftershock catalogue allowed us to better characterize the seismicity occurring during the period in which felt reports were collected and highlighted the information that is missing from operational catalogues during the early stages following a significant earthquake. The comparison between the machine-learning and operational catalogues therefore provides context for interpreting the limitations of the available macroseismic observations, as well as demonstrating the potential value of more complete earthquake catalogues for subsequent research and public information.

We have also used the enhanced catalogue as a starting point for ongoing research into the tectonic implications of the sequence. Our objective in this manuscript is to document the information available shortly after the earthquake, demonstrate how the earthquake catalogue could be improved within a relatively short period, and assess the limitations of the available felt-intensity observations. While the aftershock catalogue could be substantially improved retrospectively using the machine-learning workflow, the felt-report dataset could not be similarly enhanced. We therefore present the available intensity observations, examine their limitations, and discuss how felt-intensity data could be better collected and utilized during future events. We have clarified this distinction and the scope of the manuscript in the revised text.

My major comments are listed below.

1- The manuscript repeatedly states improved detection performance because the enhanced catalog contains more events than the reviewed catalog. However, it is difficult to assess whether this comparison alone provides a robust measure of performance. The reviewed catalog has undergone analyst quality control, whereas the enhanced catalog may contain non-earthquake sources and other erroneous detections, as acknowledged by the authors. Therefore, an increase in event counts does not necessarily imply a corresponding increase in correctly identified earthquakes.

I suggest the authors include a more thorough evaluation of the enhanced catalog. The manuscript clearly shows more detections, but I found it difficult to assess the quality of those additional events. Showing a few examples of newly detected events (maybe including a figure showing representative waveforms), comparing them with detections in the existing catalog(s), and providing 1-1 magnitude comparisons where possible would help.

I would also like to see how well the workflow recovers events already present in the reference catalog and whether there are any systematic magnitude differences. I think these comparisons would be more informative than event counts alone when evaluating detection performance.

We agree that the number of additional detections alone is not sufficient to demonstrate improved catalogue performance, particularly because the NEDB has undergone analyst review whereas the enhanced catalogue is generated automatically. We have therefore expanded the evaluation of the enhanced catalogue to provide a more thorough assessment of event recovery, catalogue agreement, magnitude differences, and location uncertainty.

For the month in which the enhanced catalogue and NEDB overlap, all 369 events in the NEDB were recovered in the enhanced catalogue within 10 s and 60 km of the NEDB origin. Of the 369 NEDB events, 170 have corresponding events in the enhanced catalogue with differential origin times <2 s and epicentral distances <5 km. Most of these matched events (92%) have magnitudes ≥3.0. These comparisons provide an assessment of how well the machine learning workflow recovers events identified through the NEDB catalogue and indicate that the enhanced catalogue is successfully recovering the higher magnitude events while also identifying additional smaller events.

We have also added an uncertainty analysis for the enhanced catalogue in the text. The overall epicentral and depth uncertainties are 6.6 km and 1.53 km, respectively. The distributions are right-skewed, reflecting a relatively small number of events with larger uncertainties and a larger number of events with smaller uncertainties. The mean difference of magnitudes between the two catalogues was 0.23, however, the analysis

showed a large distribution as opposed to a systematic magnitude bias between the catalogues.

**The results of this quality assessment are described in lines 158–178.The depth distribution is shown in Figure S5, while the uncertainty analysis is described in the text and illustrated in Figure S2-4.**

The machine-learning methods used to develop the enhanced catalogue have been evaluated in previous publications cited in the manuscript and are not the primary focus of this study. Nevertheless, we provide two examples to illustrate that the workflow is recovering arrivals associated with events in the NEDB and to demonstrate the quality of the additional detections. We provide waveform examples for the two events discussed at the end of this response. We note in the manuscript and here that the public facing NEDB is developed from analyst reviewed events that were automatically detected and located. During periods of intense seismicity, analyst review is typically prioritized for larger events. The enhanced catalogue for this event was created by the Geological Survey of Canada, a separate department within NRCan than that which is responsible for the NEDB, to fill a research and public information void.

The first example is a magnitude 3.2 event on December 7 that is not included in the NEDB, although an automatic location for the event (calculated using 12 stations) is present in the automatic, internal database. The second is an event included in both catalogues, with magnitudes of M3.0 in the NEDB and M2.8 in the enhanced catalogue. The NEDB location for this event used arrivals from 13 stations, compared with 30 stations in the enhanced catalogue. We did not perform a numerical comparison of the picks; however, visual inspection indicates that the corresponding arrivals in the two catalogues are consistent within the uncertainties of the enhanced catalogue.

While we agree that this additional information strengthens the assessment of catalogue quality, we have omitted the waveform examples from the manuscript because a comprehensive assessment of the underlying machine-learning phase-picking performance is beyond the scope of this paper. Our primary objective is to demonstrate the value of the enhanced catalogue for characterizing the aftershock sequence and to assess the additional information it provides relative to the operational catalogue.

2- This sequence is particularly challenging due to dense aftershock activity, potentially similar waveforms, and likely includes some non-earthquake signals. I would appreciate more discussion of how the workflow handles these situations and how susceptible it may be to false detections.

The manuscript notes that the enhanced catalog may contain non-earthquake sources and that additional analyses are ongoing. However, I think some evaluation of potential false detections is needed in this study. How were candidate events validated, and was any

manual review performed to confirm that detections correspond to real earthquakes? Even a rough estimate of the false-detection rate for a subset of the catalog would help put the reported increase in detections into context.

We conducted a visual quality assessment of the enhanced catalogue for 7 and 13 December 2025. All phases examined on these days were associated with real seismic events. Each day included an icequake detection, but we did not identify any landslides. We do not consider the identified icequakes to be false positives. The machine-learning workflow is designed to detect earthquake signals; however, because numerous avalanches and landslides were triggered during the aftershock sequence, signals from these sources may overlap with earthquake waveforms and potentially influence the arrival times and resulting locations and magnitudes.

During our inspection, we found that when multiple events occur within a short period of time, some adjacent events share arrivals; for example, the P-wave arrival for one event coincides with the S-wave arrival for the previous event. In the enhanced catalogue there are 143 instances in which the same arrival is associated with two adjacent events. These events are not considered false positives; however, misassociated arrivals can affect the resulting locations and magnitudes.

**These results were added in lines 166-178.**

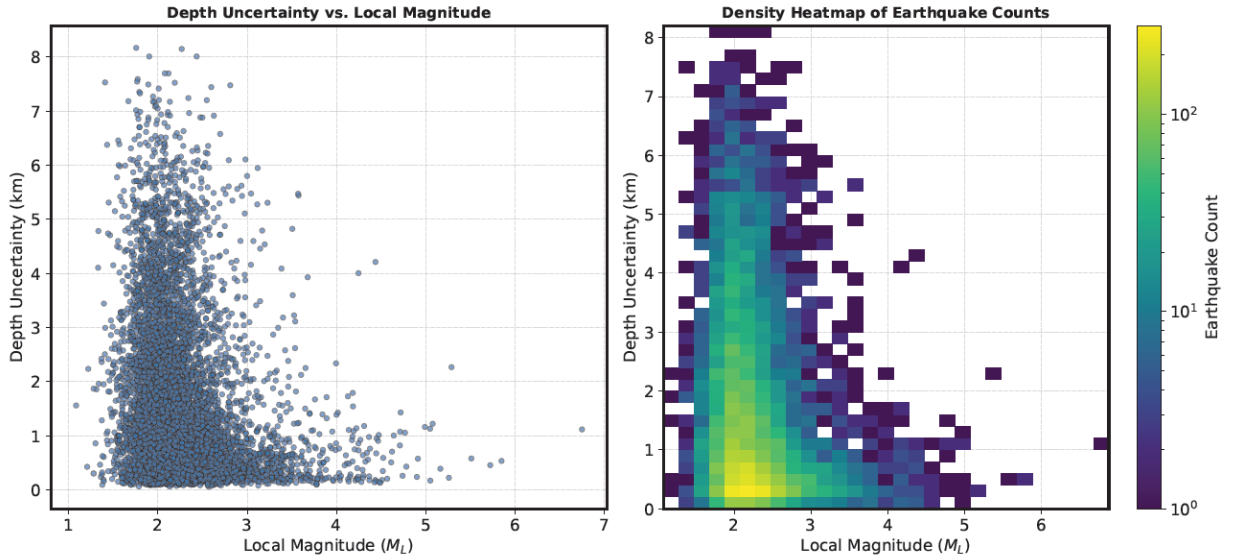
3- Lines 183–186 discuss the broad depth distribution of the catalog and note that no depths were fixed during processing. However, I don't find this discussion particularly informative without some assessment of depth accuracy and uncertainty. The key question is not whether the inversion produces a depth estimate for each event, but whether those depths are meaningful and sufficiently constrained. The manuscript also states that depth is usually poorly constrained in this region. If this is the case, then allowing depth to vary freely may simply result in a broad distribution of uncertain depths rather than more accurate hypocentral depths. In this sense, the absence of fixed depth solutions should not necessarily be viewed as an improvement over catalogs that employ fixed depths when the available data do not adequately constrain depth.

I encourage the authors to discuss how depth is constrained in the enhanced catalog and whether there is evidence that the reported depths are reliable. Are depth uncertainties available?

We agree that the main issue is whether the depth estimates are sufficiently constrained, not simply whether a depth value is reported. We have added depth-versus-distance profiles along the NW-SE aftershock trend in Figure 2 and an approximately perpendicular profile to show the **depth distribution explicitly, which is**

**included in Figure S5**.
These profiles indicate that the densest portion of the sequence is shallow, mostly within the upper ~10 km, but they do not reveal a uniquely resolved dipping structure and should be interpreted qualitatively. We therefore avoid overinterpreting individual hypocentral depths, especially for smaller events, and instead use the profiles to illustrate the first-order depth distribution of the sequence.
We have clarified in lines 209-214 that depth constraints are expected to vary with magnitude and station geometry, and that the more coherent shallow depth distribution for larger events provides stronger support for a shallow rupture than the full catalogue alone. **Depth uncertainties (mean = 1.53 km) are discussed in line 171 and shown in figure S4**

Figures of depth uncertainty distributions are provided below but are not included in the manuscript.




4- The authors note that the workflow suggested in this paper could be adopted in real time or used in a timely manner, provided that the data are available and sufficient GPU resources are available. This would be quite effective, and I would appreciate a more detailed discussion of the potential operational value of such a system. Given the challenging nature of this sequence, with its dense aftershock activity, this seems like a particularly good dataset to discuss the practical advantages and limitations of a real-time implementation.

The comparison between the two catalogues demonstrates the benefits that an operational machine-learning catalogue could provide for both research and public outreach for any situation where there are more events than what analysts can realistically review within time constraints.

The 6 December 2025 Yukon earthquake provided an opportunity to test the machine-learning workflow in a region with a lower station density than areas of higher seismic hazard in Canada that are more densely instrumented (e.g., southwestern British Columbia). The deep learning phase detectors utilized in our workflow are trained on diverse datasets, making the pipeline readily transferable to similar under-instrumented regions without requiring region-specific retraining. The automated detection of low signal-to-noise ratio arrivals remains robust even on isolated stations, though the resolution of phase association and hypocenter location will inherently depend on the station density and azimuthal coverage of the network.

**We added text to lines 142-146 that discusses transferability and limitations of the machine learning routine as well as recommendations in lines 428-434 that highlight the advantages of implementation machine learning operationally.**

The manuscript primarily demonstrates that the workflow detects more events than the reference catalog. However, it is not clear what additional information these detections provide for real-time monitoring or scientific interpretation. For example, many of the additional detections appear to be small-magnitude events occurring within a dense aftershock sequence. How would access to these events in real time improve situational awareness, hazard assessment, or operational decision-making compared to existing monitoring procedures – where a human would be looking at these events constantly? Could this lower the workload of analysts, so that they can focus on larger magnitude events, for example? If the authors envision this workflow being adopted by monitoring agencies, it would be helpful to discuss the practical benefits in a sequence like this.

Thank you for pointing out that the improvement to situational awareness and public safety was overshadowed by the enhance catalogue. The value of the enhanced catalogue beyond simply more earthquakes should be highlighted. Timely characterization of an aftershock sequence using a complete catalogue is essential not only for understanding tectonic processes, such as fault mechanisms, rupture geometry, and evolving hazards, but also for improving situational awareness among decision-makers and the public.

Current operational procedures rely heavily on analysts and are not designed to fully characterize significant aftershock sequences. Research based solely on agency databases is therefore limited by higher magnitudes of completeness and inconsistencies in analyst review procedures. A catalogue generated using consistent machine learning detection and arrival-picking methods would provide a more robust platform for research. For example, a more complete catalogue would support the identification of aftershocks in earthquake catalogues used in hazard assessment, while also providing the data needed to develop aftershock-specific hazard research.

Additionally, integrating incomplete earthquake catalogues with other datasets, such as the comparison of felt-intensity reports and aftershock occurrence shown in Figure 3, requires a more complete catalogue to support evidence-based conclusions.

Implementing machine learning methods for seismic catalogues could also reduce the operational workload, allowing analysts to focus on larger events and routine duties. This is particularly important because North America has multiple regions of high seismic hazard, creating the potential for significant earthquakes in different regions over a short period of time. Currently, detailed aftershock sequences are often left to researchers, delaying the development and distribution of complete catalogues.

A more complete catalogue would also improve public communication and emergency response. For decision makers, a robust catalogue provides a clearer picture of the extent and evolution of an aftershock sequence. In a populated area, demonstrating the thousands of aftershocks people may experience could help communicate the severity and ongoing nature of the hazard. Complete catalogues could also support aftershock forecasting for public safety, including decisions related to park closures and landslide risk. For example, Parks Canada is currently restricting backcountry access in specific areas of Kluane National Park and Reserve because of the ongoing landslide hazard associated with the aftershock sequence, based in part on analyses derived from the machine-learning catalogue.

The **Conclusions and Discussions section (lines 427-446) was modified to Conclusions and Recommendations** with a synthesized list of recommendations that captures the above response.

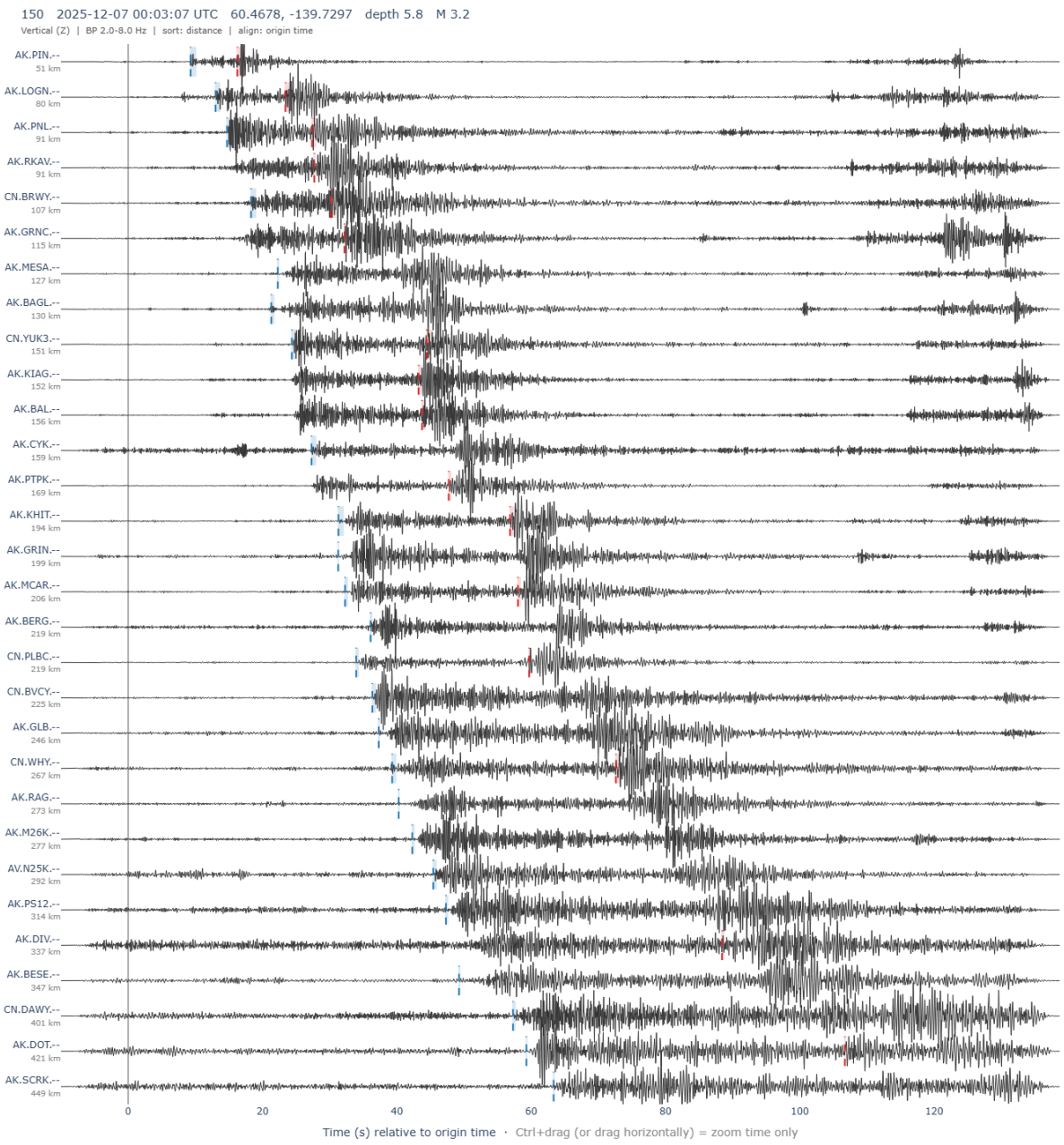



*Figure 1 - Waveform and arrival example from enhanced catalogue on December 7, 2025.*

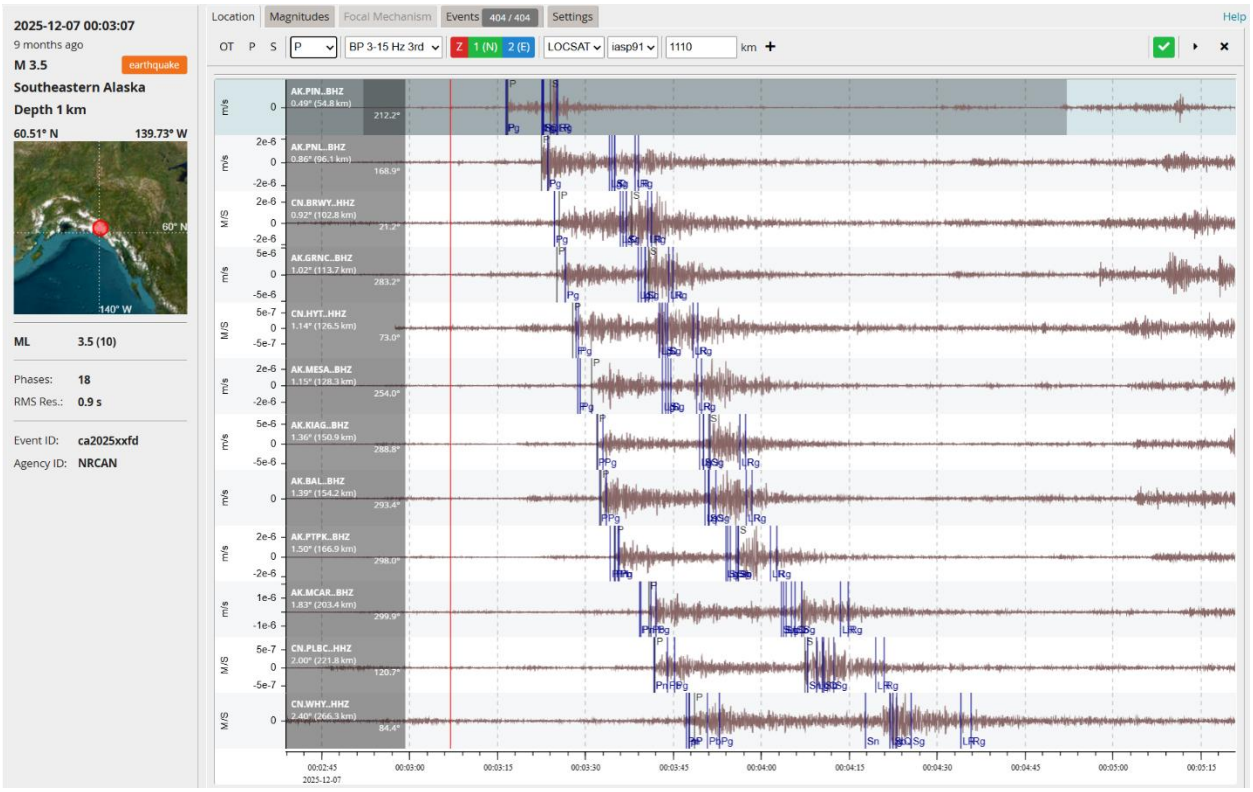



*Figure 2 - Waveform and arrival example from NEDB on December 7, 2025.*

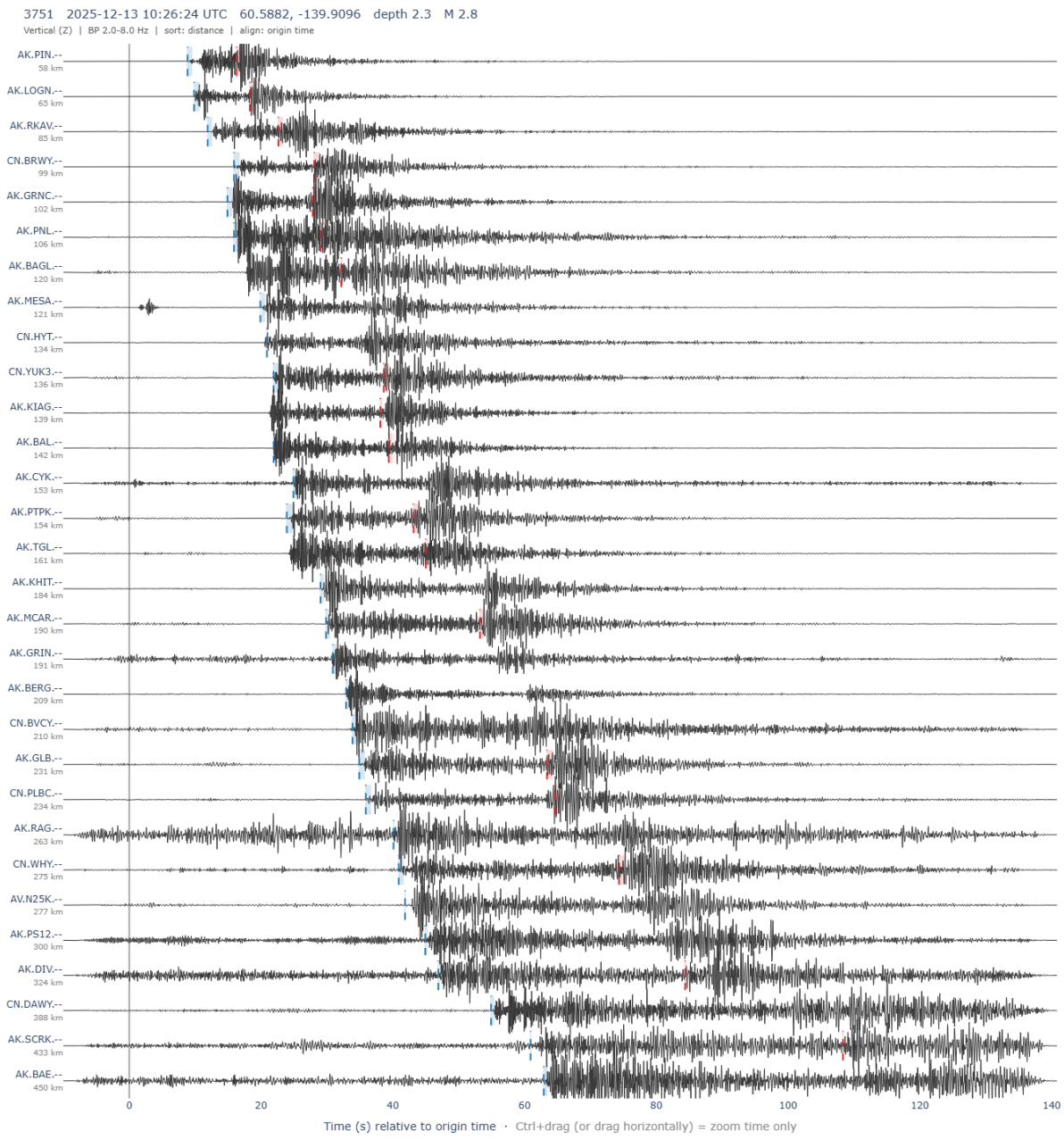



*Figure 3 - Waveform and arrival example from enhanced catalogue on December 13, 2025.*

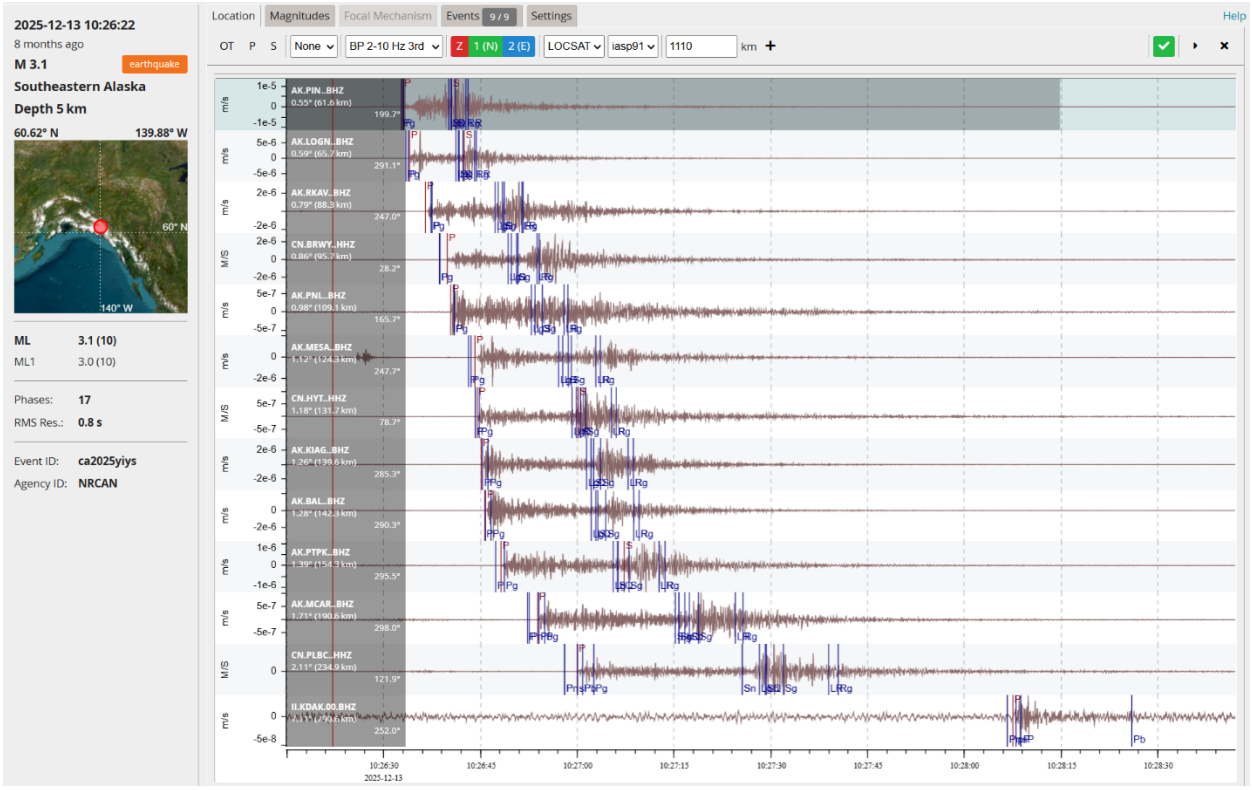



*Figure 4 - Waveform and arrival example from NEDB on December 13, 2025.*

Reviewer D:

We thank the reviewer for their thorough and constructive review of the manuscript. We appreciate the positive evaluation of the study and have addressed the comments and suggestions regarding the enhanced catalogue, comparisons with routine catalogues, and the interpretation of felt reports. Our responses to each comment are provided below.

The manuscript of Brillon et al. presents a timely assessment of the performance of routine earthquake catalogs and felt reports during the major Yukon earthquake and related aftershock sequence in a remote northern region. The combination of routine Canadian and US event catalogs, together with an enhanced catalog and the integration of felt reports, is a clear strength and delivers interesting and practical results that are relevant, also for other remote locations. The text is well written, the figures are clear, and both the capabilities and limitations of the datasets and interpretations are considered. My main comments concern, first, the development of the enhanced catalog and its comparison with routine catalogs, and second, the argumentation regarding the felt reports. Overall, I consider these comments as minor to moderate revisions that would improve the study and make the manuscript an interesting contribution to earthquake monitoring in remote regions.

**Main comments**

**Enhanced catalog and comparison with routine catalogs**

The comparison between the enhanced catalog and the routine catalogs is an important baseline of the results and data presentation in the manuscript. Did the authors cross-check the event waveforms and assess their plausibility to ensure that no noise or secondary events (e.g. landslides) were detected and falsely identified as tectonic seismic events? I miss some information on that: How many false positives were there and how were they evaluated? Was any magnitude harmonization performed between the catalogs when using the routine catalogs as a reference for the enhanced catalog? What is the location accuracy of the newly detected events in the enhanced catalog? In my opinion, the manuscript would benefit from specifying this a bit more in the main text.

While the focus of the paper was not on the machine-learning methodology itself, but rather on its application, we acknowledge that we omitted some basic information about the workflow and its performance that strengthens the results and discussion.

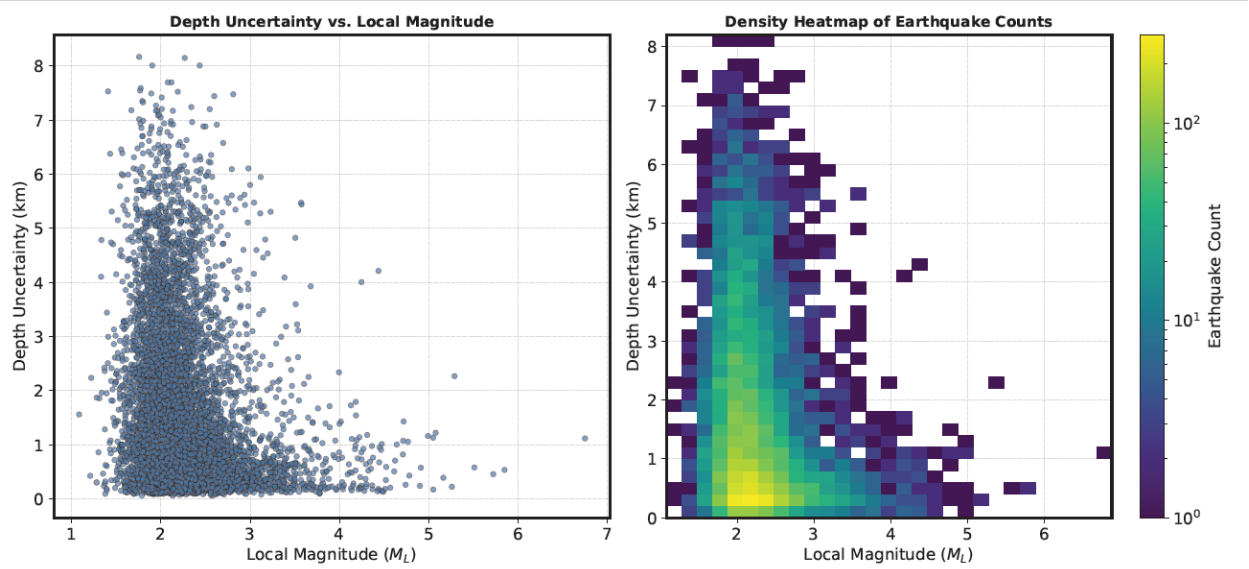
Data from 7 and 13 December 2025 were visually inspected for quality assurance. Each day had an icequake detected, but no landslides were falsely detected as an earthquake. The machine-learning routine is tailored for earthquake detection, so

landslides would not be expected to be identified as earthquakes. However, it is known that many avalanches and landslides were triggered, and their signals are likely coincident and intertwined with the earthquake signals, potentially affecting the arrival times picked by the routine and thus the final locations and magnitudes. We do not consider the icequakes false positives. **Corresponding text added in lines 166-178.**

Magnitude harmonization was not conducted. **Discussion of magnitude calculation was added to the supplement in lines 48-52**. The mean difference of magnitudes between the two catalogues was 0.23, however, the analysis showed a large distribution as opposed to a systematic magnitude bias between the catalogues. **A summary of the magnitude comparison was added in lines 159–164 and is shown in Figure S3.**

For the month in which the enhanced catalogue and NEDB overlap, all 369 events in the NEDB were recovered in the enhanced catalogue within 10 s and 60 km of the NEDB origin. Of the 369 NEDB events, 170 have corresponding events in the enhanced catalogue with differential origin times <2 s and epicentral distances <5 km. Overall, the enhanced catalogue has mean epicentral and depth uncertainties of 6.6 km and 1.53 km, respectively. The distributions are right-skewed, reflecting a relatively small number of events with larger uncertainties and many events with smaller uncertainties. **A discussion about catalogue comparisons and uncertainties was added in lines 158-178. Catalog comparisons are shown graphically in figures S2-3, uncertainties are shown in figure S4, and depth distributions are shown in figure S5.**

We include an additional figure below showing depth uncertainty versus magnitude, which, as expected, decreases with increasing magnitude.



Related to this, many figures in the supplementary material appear quite important for the main text. I understand that including them would exceed the manuscript limits, but, for example, the comparison between the enhanced and routine catalogs would be interesting to include in the main manuscript, as it clearly demonstrates the improvement in the number of detected events.

While the machine learning database is discussed in the paper, the method and assessment of the method is not the focus of the paper. We feel that moving figures from the supplement to the main text will shift the focus of the paper.

**Felt reports and their interpretation**

The felt-report analysis is very interesting and I enjoyed reading it. However, if instrumental ground-motion data are sufficient, the authors could consider including a more quantitative comparison between instrumental and reported intensity, for example by comparing an instrumental ShakeMap or PGV-based intensity proxy with the reported intensity distribution. Linked to this, the authors could test whether it is possible to average reported intensity to reduce local effects, for example due to buildings and human subjectivity, and to make the instrumental and felt-report datasets more comparable.

Seismic data within the region is limited. The weak-motion instruments, BRWY and KLRS clipped during the earthquake. BRWY, HYT and WHY have strong motion instruments, however, the HYT and WHY instruments are on mountain-tops outside of town and not representative of site conditions in town. Given this, a comparison of the instrumental intensity and reported intensity is not particularly meaningful. We have included PGA and PGV values for BRWY, HYT and WHY and note that the instrumental intensity (based on Wald et al, 1999) associated with these values is lower than what the reported intensities are. We do not compare measured PGA and PGV to the instrumental Shakemap as 1) Vs30 is largely unknown (we note this is an area that required further research) and while the Shakemap is available from the USGS based on US ground motion models, it is not available from NRCan, making the assessment of operational tools difficult. **Applicable text added in lines 383-395.**

Community Decimal Intensity (CDI) values in Figures 1 and 4 are aggregated intensity values derived from census dissemination areas, representing averages of felt reports over neighbourhood-scale areas. Figure S13 shows intensities of individual felt reports in Whitehorse. Aggregating over larger areas for figures 1 and 4 to match instrumental intensities recorded outside town is not the goal of the paper but would be an interesting exercise to embark on outside of this paper. Aggregating the data for figure S13 would decrease the ability to observe site effects, which is something we are interested in and discussed in **lines 371-382**.

We recognize that we did not consistently clarify whether plots showed aggregated or individual CDI. We have edited the captions to clarify this distinction.

Figure captions have also been modified to clarify that they are individual Community Decimal Intensity or aggregated CDI, as it may not have been clear in the previous version.

If available in the DYFI export used here, information such as building floor or building type could be valuable to better constrain and perhaps minimize this local effect on the reports. The questionnaire contains many interesting aspects, and the authors could better justify why not more of them were included. Given that the manuscript focuses on the capabilities and limitations of felt reports and how to leverage them, this would further strengthen the assessment.

We agree this is interesting information that should have been included. For this event, 64% of reports came from single-family homes, 5% from mobile homes, and 11% from apartment or condominium buildings and the rest from unclear building types, or outdoors. The highest reported shaking intensities do not align with reports from any of these building types but are distributed throughout the affected communities. **Text added to lines 349-358.**

Note that at the time of this event, the NRCAN DYFI survey did not include building floor, so we could not include that information. The Canadian DYFI survey was updated July 16, 2026, and now includes building type and floor information. While the US DYFI survey does include building type and floor, we did not have access to that information.

**Structure**

Regarding the word “assessing” in the title: The manuscript contains several important recommendation-like points (enhanced detection, inter-agency data assimilation, public communication, careful interpretation, etc.), but they are scattered; a short, synthesized set of practical recommendations (e.g. in the conclusions or end of discussion section) would help.

Related to this, since many Discussion points are already included in the text sections, the Discussion feels more like a summary of what was already discussed before. The manuscript would benefit either from reducing interpretive points in the results sections and keeping the focus on the actual report of the results, or alternatively and maybe simpler: sharpening the Discussion regarding concrete recommendations. This would improve the Discussion section, while avoiding too much repetition of points already made in the results text sections.

Thank you for the suggestion. Repeated discussion points in lines **382-415 were removed**. Discussion points in lines 382–415 that had not been addressed elsewhere were incorporated into the relevant preceding sections. The Discussion and Conclusions section was renamed Conclusions and Recommendations, and a **synthesized list of recommendations (lines 427-441) was added** to this section.

**Minor points**

Figure captions seem to be quite long, and some overlap with descriptions and interpretations in the text occurs. The authors could think of shortening or removing the interpretative points in the captions, to avoid doubling.

Captions were developed for ease of reading the supplement without the main text, however, this does have a lot of repeated information and makes the supplement very long. Captions were reviewed, and interpretive statements were removed where appropriate.

The manuscript refers to sparse seismic networks. Please shortly clarify what sparse means in this context, for example station spacing, azimuthal coverage, and detection threshold. Related to this, a short statement on the transferability of the workflow to even much more under-instrumented tectonic or volcanic regions would be useful.

In this context, "sparse" refers to regions where the seismic network contains significant spatial and/or azimuthal gaps. We have **added text in lines 80-83** discussing how sparse is defined in this region (no stations within 100 km of the epicenter). Text **added in lines 142-146** also discusses transferability and how the machine-learning pipeline can be applied to sparse networks without retraining; however, phase association and location will inherently depend on the number of stations.

Figure 1: The markers for USGS CDI and NRCan CDI are not well distinguishable. Specifically, the ones without the black edges. I suggest using e.g. rotated and non-rotated squares with edges in both cases.

Thank you for this suggestion. **Symbols were changed in figures 1 and 4** for consistency.

Figure 2: Please consider showing the event depth distribution along profiles, for example NW–SE and perpendicular to that direction.

We have added depth-versus-distance profiles along the NW-SE aftershock trend and an approximately perpendicular profile in Figure 2 to show the depth distribution now available in **figure S5.**

These profiles indicate that the densest portion of the sequence is shallow, mostly within the upper ~10 km, but they do not reveal a uniquely resolved dipping structure and should be interpreted qualitatively.
We therefore avoid overinterpreting individual hypocentral depths, especially for smaller events, and instead use the profiles to illustrate the first-order depth distribution of the sequence.

**We note (lines 207-214) in the** text that depth constraints are expected to vary with magnitude and station geometry, and that the more coherent shallow depth distribution for larger events provides stronger support for a shallow rupture than the full catalogue alone.

Figure 3c: The authors should consider using a density distribution to show how many reports occur at a specific time and distance. The crosses overlap, making it difficult to assess the actual number of reports.

We created this plot, but it is not well suited to a density plot because ~90% of reports are clustered in the 200-300 km distance range. When we plot it as a density, all the data are on top of each other. A density plot would also prevent us from scaling the symbols according to CDI. In addition, the general trend of number of events per time is already illustrated by figure 3D.

Line 98: indicate, not indicates. **Changed on line 95.**

Line 179: occur, not occurs. **Changed on line 202.**

Line 236-241: This part seems to fit better in the Discussion. **Agreed. This is also mentioned elsewhere in the paper, so these lines were deleted.**

Line 270: Do the authors mean geological site effects, building effects, or both? Please clarify here. We are referring to geological site effects, Text to clarify was **added in line 289.**

Line 392-401.: Although very interesting, this section sounds more like part of an introduction and could further motivate the study here.

The final section was restructured. These lines were deleted and the intensity information for the 2017 events was included in the **background section (lines 70-71).**

Spelling consistency: “epicentre” versus “epicenter” and “main shock” versus “mainshock.” Changed to epicentre and mainshock.

Recommendation: Revisions Required

---