Dear Editor Giuseppe Petrillo and Reviewers,

We would like to thank everyone for their helpful and constructive comments, and have included a point-by-point explanation of our responses below. Reviewers original comments are included in italic text, and our response to the reviewer in plain text. Additionally, we have an internal review from USGS (we included that here for completeness). We have highlighted text changes that were incorporated into the manuscript in light blue, and included line numbers for ease of review. The line numbers are based on the revised manuscript.

Reviewer #1 Seismica:

This manuscript is a detailed yet succinct description of the author's process for calculating moment magnitudes for small earthquakes in the Raton Basin. The authors also compare their results with other magnitude estimates for the same events. Finally they use their new magnitudes to calculate b-values for the Raton Basin and show that using moment magnitude to determine b-value leads to different results compared to using other magnitude scales (namely local magnitude). The results were very interesting and in line with many of the results I've seen in my own work which is very encouraging. I enjoyed reading this paper and I believe that it is a valuable contribution to Seismica.

Reply: Thanks for the positive comments

With that being said, I do have a few comments that I feel would contribute to the overall story of the manuscript. First, while the author does introduce and discuss other current methods of measuring moment magnitude for small earthquakes, this discussion doesn't take place until the very end of the paper. As I was reading, I often found myself thinking "why did the authors choose this method over other very similar methods like Mayeda's coda-envelope method which is very similar?" I think that weaving these explanations through both the introduction and the discussion will help to motivate and present this work as a complete story rather than just reporting results for a particular method/study region.

Reply: There are several methods to obtain source parameters (or moment magnitudes) . Which one is better is a whole different research question but we prefer this method over others such as single spectrum as it is less affected by radiation patterns and we have a relatively small number of seismic stations. Additionally, the CCT method (Mayeda's coda-envelope method) is a data-based approach that requires some empirical calibrations while Qopen is based on a physical model. We added a brief introduction to other methods earlier in the manuscript and highlighted a couple of other relevant points you asked in other sections below.

- L66-L76: For smaller earthquakes, where full moment tensor inversions are often unfeasible, other techniques have been developed to estimate M_W or other source parameters. These techniques involve measurements of Ω0 including individual spectral fitting of body waves, spectral decomposition, generalized inversion techniques (Shible et al., 2022; Morasca et al., 2025), or coda wave analysis (Mayeda et al., 2003; Eulenfeld et al., 2022), amongst others. Such techniques had been applied successfully to both natural and induced sequences, such as the 2016-2017 Central Italy (Morasca et al., 2022; Kemna et al., 2021), the 2019 Ridgecrest (Trugman, 2020; Mayeda et al., 2025) earthquake sequences, as well as seismicity associated with wastewater injection (Kemna et al., 2020; Shelly et al., 2022), hydraulic fracturing (Wang et al., 2020a), and enhanced geothermal systems (Eulenfeld et al., 2023). These studies provide essential insights into earthquake size, stress-drop scaling, source heterogeneity, and rupture complexity. Understanding these source parameters (including M_W) is a critical step toward improving hazard assessments and mitigating the risks posed by induced seismicity.

- L132-L136: We choose the Qopen method over others because coda-based approaches are less sensitive to radiation pattern effects and because we have a relatively small number of seismic stations. Additionally, the Qopen method is based on a physical model and does not require calibration or information from previous events.

Also, I am quite concerned about the fact that only 1/3 of the catalog is recalculated. I recognize that providing moment magnitudes for these small events is a very difficult task and sometimes there is no choice to exclude many events when the data just isn't good enough. However, I feel that this is a bit "glossed over" in the text. The authors do correctly compare b-value only for overlapping events but if this is part of a larger discussion of how these different magnitude scales affect our interpretation of seismic hazard, then we need to also discuss the completeness limitations for these methods as well. Please see some more specific comments below.

Reply: We agree that losing ⅔ of the catalog is a large loss but the number of events (n~ 30,000) we have M_W is large in comparison with prior studies that report b-values (Shelly et al., 2022; Baltay and Abercrombie 2025; Ross et al., 2016) which contains n < 10,000. Our study is a substantial progress in the direction of scaling up use of M_W to larger catalog with lower M_C.

However, the number of samples (n=31,581) we have to estimate the magnitude of completeness and b-values is still large (n > 10,000) and should not generate any major biases. Ogata and Yamashina (1986) and Shi and Bolt (1982) have shown that b-value biases are proportional to 1/n, which in this case is minimal.

We compare the magnitude-frequency distributions, as you requested below, of the whole catalog (n=95,993) and the subset of events (n=31,581) with a local magnitude. This is shown in Figure S4. The subset is mostly incomplete, in comparison with the whole catalog, for small events, i.e., $M \leq 0$, which is well below the M_C for the subset or for the whole catalog. Moreover, both distributions are highly similar above M_C and the b-values derived from them are equivalent, which confirms that our subset completely captures the population relevant for a reliable b-value comparison between the two magnitude scales (M_W and M_L).

Regarding seismic hazard analysis, it is the larger earthquakes that are often of greater concern and a catalog of ~30,000 Mw still provides a strong foundation for hazard analysis. Moreover, seismic hazard analyses are limited by the catalogs used because they are commonly declustered or culled with a higher magnitude of completeness (e.g., >3).

- L370-L381: In the Raton Basin, previous studies (Glasgow et al., 2021; Jamalreyhani et al., 2025) have reported MC ≈ 0 and b ≈ 1, based on larger catalogs with local magnitudes (with n ≈38,000 and n ≈ 96,000 events, respectively). For our analysis, we used a subset of 31,581 events for which both M_L^{SJ25} and M_W^{Qopen} magnitudes are available. Although this is smaller than the full catalog (n =95,993), a comparison of their frequency-magnitude distributions confirms that this subset does not bias our results. As shown in Figure S4, significant incompleteness in the subset occurs only for magnitudes $M \leq 0$, which is well below the M_C for both the subset and the full catalog. Furthermore, the distributions are highly similar above M_C , and the b-values derived from them are equivalent. It is uncommon to have such a large number of events (n =31,581) with M_W available for b-value estimation as we have here; prior studies with moment magnitudes (Shelly et al., 2022; Ross et al., 2016; Baltay and Abercrombie, 2025) have used far fewer events (n<10,000). This is a significant advantage, as seismic hazard analysis (Valensise et al., 2024; Taroni and Akinci, 2021; Teng and Baker, 2019) are often limited by catalogs that are either declustered or culled with a high M_C , i.e., $M_C \geq 3$.

Line Comments

Title - Suggest saying "frequency-magnitude" instead of frequency magnitude"

Reply: We changed the title to your suggestion

Line 24: Perhaps it would be helpful to label the local magnitudes in these cases with a specific label to differentiate (i.e. ML,USGS or something similar as you have done later in the paper). Also what is the source of the ML estimates in the first equation. Is it all non-USGS MLs that you have collected?

Reply: We believe it is not necessary to clarify upfront that the local magnitude is not from the USGS. We introduce all of that detailed information later in the text.

Line 45: Suggest changing "notice" to "observed"

Reply: We changed the verb to your suggestion

Lines 53 - 55: I think the sentence "Despite such widespread use ... variable attenuation structure" should have some references. Shelly et al. 2022 would be good to include here which you have referenced later on in the paper as well as some others.

Reply: We added to this sentence the reference of Deichmann 2006.

Lines 62 - 65: The transition between the end of this paragraph and the next one is quite abrupt and sounds a little bit awkward. The author discusses how moment magnitudes are more useful because they are physics-based and mentions that they often cannot be measured because of limited signal-to-noise ratio. I would suggest including here a very short summary of some other current methods and a more clear transition to the topic of induced seismicity and the need for better magnitude estimates. Alternatively the author could include subheadings to distinguish the two sections.

Reply: See reply to major comment #1 above

Line 68: "between 2001 and 2025". When I read this I assume this means the beginning of 2001 to the beginning of 2025 as 2025 is not finished yet. Is this what you meant? Perhaps consider adding months (i.e. January 2001 - January 2025)

Reply: Correct. We added the months for clarity.

Line 72: From the abstract and introduction it is clear that the motivation of this study is to provide moment magnitude estimates for this region of induced seismicity, but it is less clear why this method will be an improvement over other methods. You discuss this a bit near the end of the paper but I feel that there should be a short summary in the introduction along with my comments for lines 62-65.

Reply: We address this comment in L132-134 We chose the Qopen method over others because coda-based approaches are less sensitive to radiation pattern effects, its open access software has been tested in recent publications (Eulenfeld et al., 2023; Eken, 2019; Eulenfeld and Wegler, 2017), and it is well-suited to a small network of stations like we have for the Raton Basin.

Line 94: This is more for my personal curiosity and may not really need to be addressed in the manuscript but are the eight stations that you used all that you had available, or did you specifically choose stations so that they had ~30 km spacing? If the latter, why 30 km specifically?

Reply: We use these eight stations because they are all of the available stations in the Raton Basin. Some of the stations are recording data in campaign mode as well. We added the following sentence L107-108: The events were detected using all eight of the publicly available broadband stations in the Raton Basin, which provides about 30 km spacing.

Is there another reference for the training dataset or was this part of Zhu and Beroza's (2019) work? I'm also curious how training PhaseNet on picks from Northern California will perform for picking waveforms in the Raton Basin. Again, this is more of a side comment. Feel free to include an answer to this question as you see fit in the manuscript.

Reply: This was part of Zhu and Beroza's work. The training dataset was recompiled by them as well but it comes from the Northern California Earthquake Data Center Catalog. Glasgow et al., 2021 did a comparison between 50,000 manual picks and PhaseNet picks (their Figure 3) in the Raton Basin (same stations as we used in this study). Glasgow et al., (2021) mentions that 98% of P and 94% of S phase picks are within one tenth of a second of each other. We added this to the manuscript L111-112: Previously, Glasgow et al., (2021) showed that PhaseNet performs well in the Raton Basin, detecting phase picks within one tenth of a second in comparison with 50,000 manual picks.

Line 105 - 107: How do these two methods scale with each other? Do they give the same answer for the same event?

Reply: The differences between the M_Ws of these two methods are on average 0.08 indicating that they are consistent with each other. We added the following sentence L125:Both methods provide similar values of M_W for the same event with average differences of 0.08.

Line 218: Losing 2/3 of the catalog is quite significant but the author doesn't really discuss this in any more detail. How does this affect the resulting b-value? Is there a particular magnitude range or location that this incompleteness affects the most? Is it still helpful for seismic hazard analysis to only have 1/3 of the catalog? Can anything be done in the future to improve this?

Reply: We replied to this topic of feedback in detail below Reviewer 1's major comment three.

Line 258: I think "MW and MW" should read "MW and ML"

Reply: You are correct. We apologize for the typo. It is fixed now.

Figure 3: It is quite difficult to differentiate between the crosses and stars as well as the two colors that you chose. Perhaps it would be helpful to separate the left panel into different panels or change the color/symbols? For the comparison between your results and ML,SJ25 I would suggest a 2D density plot rather than a scatter plot. This would show the relationship more clearly (use a log scale for the color if it doesn't work out at first. I've had to plot very similar figures). It appears that there may be two distinct groups of events divided around Mw = 1 but I'm not sure. Also, it looks like there is more data below Mother ~ -0.5 but the graph looks cut off. Why do we not see the smallest events? Last thing, I may have missed an explanation in the text but it looks like there is a strict lower limit of ~0.5 for the Mw,Qopen method. What is causing this?

Reply: We modify Figure 3 so it is easier to differentiate where the majority of our data is. We adopt your suggestion of changing the figure. Now, some of them are density plots. Figure 3a and 3b contains density plots showing the relationship between M_W^{Qopen} and M_L^{SJ25}, and M_W^{Qopen} and M_L^{CC}, respectively. Figure 3c contains the other magnitudes we analyzed in our study (i.e., we adopted your suggestion of separating the panels). There were only a few data points at $M_L \sim -1.0$, we increase the limit axis so it covers the entire magnitude range that is shown in Table 1. We did not state that there is a lower limit of 0.5 for Qopen and we did not impose any lower limit in the study. The lowest magnitude we were able to obtain with Qopen was a 0.42 (see Table 1 for the full range).

Figure 6: Maybe this is just a personal preference but I think that the results for the Mc test could be moved to the supplementary material. The results for Mc do affect the resulting b-values but it doesn't seem to be a focus of your major discussion points.

Reply: This is a key result of our manuscript since we don't know which is the best Mc, and we show that the b-values remain on a similar range for the obtained Mcs. Additionally, we added another method for Mc as requested for reviewer #2 and added more discussion regarding biases on Mc. See L382-394. So, we prefer to keep this topic in the main text.

Line 333-334: How does the incompleteness of your magnitude method affect the b value results? (See comments for line 213) I understand that by calculating b value only for events that have both magnitude types essentially shows that the difference in b value is in fact due to the differing magnitude types. However, I feel that a discussion of how incompleteness due to the magnitude type is necessary for a reader to fully understand the whole picture, particularly in reference to seismic hazard.

Reply: We addressed this topic above, along with your comments in L370-381.

Here, we have evidence that differences in b-value are caused by differing magnitude types. However, we can't really report these b-value results as the "correct" b-value either because 2/3 of the catalog is missing ,unless you can show that the incompleteness is not affecting the b-value. This can be investigated easily by plotting the magnitude frequency distribution of the original ML estimates and then the original ML estimates for only the events you were able to recalculate. Where they diverge will show you what proportion of events are missing.

Reply: We addressed this topic above, along with your comments in L370-381 and added the suggested figure in the supplemental material (Figure S4)

Lines 336 - 338: This sentence is stated in a very similar way earlier in the paper.

Reply: We removed this sentence

Line 407: Is "felt2" a typo?

Reply: It is not a typo. There is a footnote related to these felt events.

Lines 448: "The CCT method… optimal inversion scheme on each dataset: This is very valuable context about the other current methods that also provide similar results. I think it would be useful to have a condensed version of this introduced earlier in the paper (perhaps in the introduction near lines 62-65). As it is written now, the reader does not have any context about this method in relation to any other similar Mw measurement methods until the very end of the discussion.

Reply: We addressed this point in L66-76

Reviewer #2 Seismica

This study explores the temporal and spatial variability of the Gutenberg–Richter b-value in the Raton Basin, with implications for fluid-injection-induced seismicity. The paper is clear, well organized, and employs multiple b-value estimation techniques. This methodological diversity is commendable and enhances robustness.

Reply: Thanks for your feedback on the manuscript. We note that our study focuses on estimation of moment magnitudes and how it affects other earthquake statistics including b-values calculated by multiple methods. We do not aim to analyze spatiotemporal variability linked to fluid injection.

Several important issues concerning catalog completeness, bias correction, and interpretation of b-positive outcomes must be addressed before publication.

MAJOR COMMENTS
1) The authors apply three estimators (MLE, b-positive, and K-MSlope), which is a strong methodological choice. However, the discussion of their relative performance and theoretical underpinnings is insufficient. The b-positive estimator (van der Elst, 2021) was originally designed to mitigate the effects of transient incompleteness in aftershock sequences. Yet, as shown by Lippiello and Petrillo 2024, the b-positive estimator itself can exhibit small biases. The b-more-positive variant extends this approach by using all possible magnitude pairs (not only consecutive ones), thereby increasing efficiency and removing residual bias.

Therefore, I recommend that the authors to explicitly discuss how b-more-positive could alter or confirm their results and clarify whether their b-positive implementation includes filtering conditions to ensure unbiasedness.

Reply: We implemented the b-more-positive method as described in Lippiello and Petrillo, 2024 and it is consistent with our results from other methods. We added in Figure 4a and 4b the positive magnitude pairs and the respective b_{++} fit. Our results remain similar: b values are different depending on whether the input is local magnitude or moment magnitude. b_{++} is 1 for M_L^{SJ25} and is 1.3 for M_W^{Qopen}.

We had mentioned the preferred solution for b_+ with the condition of $M_C` \geq 0.2$ (See L236). Even with other values of $M_C`$ (Figure 5c), the results for the b-positive are similar. We also include in Figure 5d the dependence of b_{++} as a function of M_{th} but the results are similar to the b_+ method.

We added the following lines regarding this comment:

- L238-243: (iii) the b-more-positive (b++) method (Lippiello and Petrillo, 2024) which is an extension of the b+ method that include all possible magnitude pairs (not only successive ones) and is defined as: $b_{++} = \frac{\log_{10}(e)}{(\overline{|M`|_l} - M_{th})}$, where $\overline{|M`|_l}$ denotes the average magnitude difference for all $\Delta M = m_{i+l} - m_i$, where l ≥ 1. M_{th} is a threshold value. We show that different M_{th} values lead to similar conclusions. We use a value of l = 10, as suggested by Lippiello and Petrillo (2024).
- L312-319: The b-positive (van der Elst, 2021) and b-more-positive (Lippiello and Petrillo, 2024) methods are more robust in comparison with the other methods we applied in our study. However, the minimum magnitude difference (M_C) parameter or the minimum threshold parameters (M_{th}) and how it affects the b-value estimation has been investigated less. Figure 5c) and 5d) shows the dependency of b+ versus M'_C and b++ versus Mth using M_L^{SJ25} and M_W^{Qopen}, respectively, which shows that there is some minor variability on the b-value depending on those two parameters. For example, applying the b-positive method to M_L^{SJ25} with a M′C = 0.2 leads to b = 0.96 while using a M′C = 0.5 results in b = 1.03. Nevertheless, we note that b+ and b++ values are generally close to 1 when we apply it to M_L^{SJ25} while b+ and b++ values are overall larger than 1.1 for M_W^{Qopen}.
- L418-422: Some of the most robust methods for estimating b-values are the b+ method (van der Elst, 2021) and the b++ method (Lippiello and Petrillo, 2024) because they are relatively insensitive to transient changes in completeness, such as short-term aftershock incompleteness or evolving seismic network configurations. We still observe that changing the magnitude scale from M_L^{SJ25} to M_W^{Qopen} results in an increase in the b+ from 0.96 to 1.25 and for the b++ from 1.0 to 1.3

2) The authors estimate Mc using three traditional methods: Maximum Curvature (MAXC), Goodness-of-Fit (GFT), and b-value Stability (MBS) and prefer MBS as the reference. While this is acceptable, these approaches are known to suffer from under- or overestimation in heterogeneous or time-varying catalogs .

More robust alternatives exist:

Coefficient of Variation (CV) techniques (Godano et al., 2024) provide consistent estimates with quantified uncertainty.

Verification-based approaches (Petrillo and Zhuang, 2023) can detect time-dependent completeness variations, a crucial factor in induced seismicity studies.

Reply: Thank you for your suggestion. We added the method by Godano et al., 2024 to the text, added its result to the main text and changed our preferred M_C solution for this one. We modified Figure 4 to show the results based on this M_C. Since with the addition of this method there are too many M_C, we added a table with all the M_C we obtained for M_W^{Qopen} and M_L^{SJ25}. Our results or conclusions do not change based on the new preferred MC using Godano's method. We put the lines added or modified below:

- L214-222: Multiple methods have been proposed to obtain MC. We use the following approaches to determine MC : (i) the Maximum Curvature (MaxC) technique (Wiemer and Wyss, 2000), (ii) the Goodness of Fit Test (GFT; Wiemer and Wyss (2000)), (iii) M_C by b-value stability (MBS; Cao and Gao (2002)) and (iv) M_C by coefficient of variation techniques (MCV) (Godano et al., 2024) that can be considered a generalization of Cao and Gao (2002). For the M_C obtained with the MaxC technique, we include a correction of +0.2 as suggested by Woessner and Wiemer (2005). Additionally, we define the acceptable M_C range as the minimum and maximum values from these methods. Given that the choice of M_C directly impacts the b-value, we focus on the M_C obtained by Godano et al. (2024) method since the other methods might generate overestimations or underestimations of the M_C (Woessner and Wiemer, 2005), although we confirm that our conclusions hold across all methods.
- L303: Table 2 shows the M_C results for all the methods, which ranges between 1.2-1.4 for M QopenW and 0.5-0.7 for M SJ25 L . The M_C values, obtained using the Godano et al. (2024) method, are 0.5 for M_L^{SJ25} and 1.2 for M_W^{Qopen}.
- The b-values obtained with the new M_C are shown in the legend of Figure 4 which are generally close to 1 for local magnitude and > 1.2 for moment magnitude.

3) The correlation between b-value fluctuations and injection activity is plausible but remains qualitative. The paper would benefit from a quantitative correlation (e.g., Pearson r) between injection rates and b-values and confidence intervals or bootstrapped uncertainty for all estimators.

Reply: In this study we do not aim to relate the spatio-temporal variations of the b-value with the injection activity. Our goal is to evaluate how different magnitude scales (especially Mw versus M_L) and b-value estimation methods estimate may change the basic earthquake statistics derived from a high-resolution catalog. There is no analysis of time-dependent b-values in the manuscript and we prefer to maintain that choice in the design of our study. . We understand that other studies (van der Elst, 2021; Lippiello and Petrillo, 2024; and others) in other regions have focused on that variability and acknowledge that it could be valuable to explore in future studies of the Raton Basin. We refer the reader and reviewer to some previous studies in the Raton Basin if they are interested in some relationships between the seismicity and the injection rates (Nakai et al., 2017; Glasgow et al., 2021; Stokes et al., 2023).

L83-85: "Previous studies (Nakai et al., 2017; Glasgow et al., 2021; Wang et al., 2020b; Stokes et al., 2023; Jamalreyhani et al., 2025) have identified that the seismicity is hosted by basement rooted fault segments and studied the spatiotemporal evolution of the seismicity and wastewater injection."

The study is valuable and methodologically diverse but requires a deeper discussion of estimator biases, completeness reliability, and the implications of modern developments in b-value estimation. Incorporating recent unbiased methods would substantially strengthen the paper credibility and impact.

Reply: Thank you for the comment. We added the two methods you suggested are more robust for the magnitude of completeness (Godano et al., 2024)and for the b-value estimation (Lippiello et al., 2024b). Our conclusions remain similar even with the most robust methodologies.

We added more discussion regarding the magnitude of completeness and some of their issues

- L382-394:Numerous methods exist for estimating M_C (Wiemer and Wyss, 2000; Cao and Gao, 2002; Godano et al., 2024b; Lombardi, 2021; Godano, 2017; Wang et al., 2025; Amorese, 2007; Ogata and Katsura, 1993; Taroni, 2023). In this study, we tested four of these methods (see Table 2) on events with M_L^{SJ25} andM_W^{Qopen}, which yielded M_C ranging between 0.5–0.7 and 1.2–1.4, respectively. This dependence of M_C on the magnitude type is perhaps expected, as M_W values are systematically higher than M_L values, naturally leading to a higher M_C when M_W is used. The accuracy of any M_C estimate is inherently dependent on the number of samples and the shape of the magnitude-frequency distribution. Furthermore, the methods themselves have known biases: the MBS method can overestimate MC (Lombardi, 2021; Zhou et al., 2018), while the MAXC method tends to underestimate it (Mignan and Woessner, 2012; Zhou et al., 2018; Woessner and Wiemer, 2005), often adding an arbitrary +0.2 correction. The GFT method's performance is variable, potentially leading to either underestimation or overestimation depending on the dataset (Woessner and Wiemer, 2005; Zhou et al., 2018). Although the MCV method has been tested on various regions (California, Italy, New Zealand, and Japan) and synthetic catalogs (Godano et al., 2024b,a), its performance in other tectonic settings requires further validation.

Internal reviewer USGS

Review of Peña Castro et al., USGS internal review

David Shelly, 8/25/2025

This study estimates moment magnitudes for a large catalog of earthquakes in the Raton Basin and analyzes the resulting statistics and implications. This study is very careful in its magnitude determinations and looks at various approaches to b-values. Overall, this paper is good shape. I give some specific comments below to help improve the manuscript, mostly by clarifying the text in various places. Included in these comments is the argument that ML doesn't "underestimate" earthquake size – it's simply a different measure of size than Mw.

Specific comments

Line 59: I don't think it's accurate to say that seismic potency relies on estimation of seismic moment. Potency is just average slip times area, without the rigidity factor required to compute moment. So I suppose you could instead say that both rely on estimation of potency.

Reply: Thanks for the comment. We apologize for the confusion. We change the term to potency magnitude (M_P) to be more clear. According with Trugman and Ben-Zion (2024): $M_P = \frac{2}{3}(\log_{10}P_0 + 5.4563)$, with $\log_{10}P_0 = \log_{10}M_0 - \log_{10}\mu - 11$ and μ corresponds to rigidity. We correct the sentence to L59-61: Moment magnitude (M_W ; Hanks and Kanamori (1979); Shearer (2019)) or potency magnitude (M_P ; Ben-Zion (2001); Trugman and Ben-Zion (2024)) offers more physical-based measurements of earthquake sizes, both relying on the estimation of the seismic moment (M_0).

Line 63: low frequencies used for moment tensors are usually <0.1 or <0.2 Hz (10 s or 5 s period). M<3 earthquakes generate very little energy at these low frequencies, so some studies have come up with ways to use higher frequencies. There is plenty of energy at <10 Hz for these smaller events, so that's not the limitation. Please clarify text.

Reply: We were referring to the limitations of moment tensor inversions that only use frequencies lower than 1 Hz. We added the following sentence L63-65: Moment magnitudes are commonly reported for moderate to large earthquakes (M > 4) but less often reported for smaller earthquakes (M < 3; Alvizuri and Tape (2016)) due to the limited signal-to-noise ratio at low frequencies, i.e., <1 Hz, that are necessary for waveform inversions that involve full moment tensors analysis.

Line 64: Several studies have suggested that some induced seismicity existed well before 2009 (indeed, the current study goes back to 2001), so probably replace "occur" with "increase" (or similar).

Reply: We changed the verb to your suggestion.

Line 102: It would be good to mention that the amplitude is measured here as a 3c vector, and also that a Wood-Anderson response is applied (if I'm understanding Glasgow et al 2021). Is the amplitude a ground motion amplitude or an instrument amplitude (the latter including the instrument gain)? ML was originally defined as the instrument (Wood-Anderson) amplitude, but more recent implementations have deviated from this.

Reply: We added the information you requested regarding this calculation L116-120: The local magnitudes were estimated using the equation from Glasgow et al. (2021): $M_L = \log_{10}A + 2.56\log_{10}D - 4.6$, where A represents the maximum three-component vector amplitude (in micrometer) and D is the distance (in km). The 3-component waveforms were filtered from 0.01 to 40 Hz, corrected for instrument response, and converted to Wood-Anderson displacement.

Section 2.2: I appreciate including a brief summary of the method. Could you include a brief discussion in the early part of this section about what kinds of datasets are a good fit for this method? For example, it sounds like it's an inversion, so presumably it works best with large numbers of events? Do events need to be within a similar depth range to work well (like for the Mayeda method)?

Reply: Thanks for the comment. We added some of the regions where Qopen has been used which ranges from local events to more regional distances. It does require some number of events but we are not aware of an exact number to have better results. For example Eulenfeld et al., (2022) only used 39 earthquakes for the attenuation estimation and site-amplification factors, while Sens-Schönfelder and Wegler, (2006) only analyzed 9 earthquakes (but these were large events M>4), Eulenfeld et al., (2023) only used 36 so it seems that it does not require a relatively large number of events (n <50). In our case, we tested first with the catalog of Glasgow et al., (2021) which contains fewer events over only four years of data and then we got similar results when we used the more complete catalog of Jamalreyhani et al., (2025). Moreover, we used 151 events in the first steps which should be more than sufficient in our study. Most of the events (92%) in our catalog are located at depths between 5-10 km so we can't assess how well the method might perform in extreme cases where there is shallow seismicity (< 3km) and deeper seismicity (5-15km) as the case shown by Patton et al., (2025) but if the attenuation conditions are significantly different between the shallow and deeper layers it might require the inversion to be performed as separate sets.

We added the following lines L136-141: The method has been applied to investigate source parameters at multiple regions with different network geometries including: the community stress drop validation study of the 2019 Ridgecrest earthquake sequence (Baltay et al., 2024), earthquakes recorded by a local network in an enhanced geothermal system in Helsinski (Eulenfeld et al., 2023), earthquakes in the Anatolian Fault Zone (Turkey) (Eken, 2019; Izgi et al., 2020), an earthquake swarm near the Czech–German border region. An earlier version of the method was used to investigate regional earthquakes in Germany (Sens-Schönfelder and Wegler, 2006).

Line 179: Shelly et al, 2016 argued that depending on the degree of fluid pressure confinement, earthquakes may preferentially occur along the same larger fault structure (low b-value) or may randomly sample faults in the crust (higher b-value). So you might say something like "...attributed to interactions between fault structure and pore pressure..."

Reply: We added your suggestion. L207: Such variations might be attributed to interactions between fault structures and pore-pressure (Shelly et al., 2016),...

Line 214 and elsewhere: there seems to be an odd space after the comma in numbers throughout, e.g., 95, 993 rather than 95,933.

Reply: We fix the space. It was automatically generated when we included the numbers using $$ in latex.

Figure 3a: it's hard to distinguish the different symbols here, because they're largely on top of each other. At a minimum, it would be helpful to make the symbols larger in the legend (also for Fig 4a).

Reply: We changed Figure 3a and also Figure 4a as there was too much overlap between the symbols. Reviewer #1 suggested a density plot so we changed the Figures to that.

Figure 3b (and lines 238-240): This plot is perhaps a bit confusing/misleading, because delta-M strongly depends on the magnitude range examined, and different magnitude have different ranges. Please make the labels larger at the bottom of the plot (M super and subscripts are very difficult to read).

Reply: Figure 3b in the old version is now Figure 3d. We increase the font sizes of the super and subscripts. The box-plots are just to provide a simple visual metric of the magnitude of variations between different magnitude types.

Line 255-256 (and several other places in the manuscript): I don't think it makes sense to say that "...ML underestimates the true size of earthquakes..." ML is one measure of size, while Mw is another. It's almost like this text is trying to say that if you take ML and assume it's Mw and then try to estimate moment, you will underestimate it. If that's what's intended, then it should be described more clearly.

Reply: We agree that M_L and M_W are measurements of the size of an earthquake. However, if these two variables are measuring the same phenomena then they should always give, roughly, the same number. Several studies have shown that M_L and M_W do not converge to the same numbers for M < 3. Hanks and Boore, (1984), Deichmann, (2006), Baltay et al., (2017) and others have shown the theoretical reasons why they do not converge at the smaller magnitudes and Abercrombie (1996), Ross et al. (2016), Deichmann (2017), and others observe such difference between M_L and M_W for different regions.

We change the wording in this sentence to L345-347: These linear relationships suggest that M_L may significantly underestimate the true size of small earthquakes in the Raton Basin, assuming M_W is a more reliable measure of earthquake size.

Lines 262-263 and Figure 4b: The text says that the regression between Mw and Mw_Qopen was obtained by ODR. Was a slope =1.00 obtained from this method? Visually it appears that the slope in Figure 4b should be slightly steeper to minimize orthogonal residuals. Please clarify.

Reply: Yes, the slope was obtained with an ODR. Figure 4b now is Figure 3c. In such a Figure we added a 1:1 line and it is clear by comparing the suggested regression (red dashed line) and the 1:1 line that they have the same slope.

Fig 4 caption, last line: I think it's helpful to distinguish between individual magnitude uncertainty (typically 0.1-0.2) and overall bias (systematic differences, in this case the 0.09 offset).

Reply: We added the following in the Figure 3 caption: The RMSE and overall bias (offset) of 0.09 is below 0.1-0.2 which are typical errors reported for an earthquake magnitude.

*In the b-value discussion, it may be worth noting that b-value scales directly with the slope of the magnitude conversion relation. So if Mw=0.7ML+0.96, and b-value for ML=1.0, then b-value for Mw is 1.0/0.7=1.43. Likewise, if Mc for ML=0.7, then Mc for Mw=0.7*0.7+0.96=1.45. Now, I understand that the equation is just a regression, and that you're not using it to determine Mw, but this still could be helpful context to understand the differences you're finding in these values for Mw and ML.*

Reply: You are correct that the b-value and the magnitude of completeness (Mc) scale roughly with the slope of the magnitude conversion relationship. However, there is already a lot of discussion and existing issues related to estimating Mc and b-values using different methods, which often lead to underestimations or overestimations. Therefore, we would rather not add more confusion to this topic.

Lines 362-363: The problem with computing b-values by least squares isn't underestimation (is this confused with using least-squares to compute magnitude conversions?), but rather that least squares weights points at high magnitudes, which have very few contributing events, the same as at low magnitudes, with many events. In contrast, maximum likelihood weights all events equally. Please clarify text.

Reply: We added your point L412-415: The b_{LF} method is generally discouraged due to its tendency to underestimate b-values (Castellaro et al., 2006; Sandri and Marzocchi, 2007; Geffers et al., 2022), as it weights high-magnitude events (small samples) the same as low magnitude events (large samples).

Some parts of the manuscript seem slightly repetitive, so it may be possible to compress certain sections without loss of information. For example, the first 2 paragraphs of section 4.2 (lines 333-349) seem to largely contain discussion similar to that earlier in the manuscript.

Reply: We removed a couple of sentences that are a bit repetitive in the text.

Line 383: "...inaccurate b-values..." b-values are different when computed using ML or Mw, but there is no "true" b-value. So I don't think you can say that one is more accurate than another. b-value is just a statistical parameter. Now, it might be true that one is better than other in terms of extrapolating to rates of larger earthquakes, but that's a separate question.

Reply: We changed the description to different b-values. We agree with your point that one might be better than other for extrapolating to rates of large earthquakes but that's beyond our study L435-436: Therefore, we expect that other seismically active regions might also have different b-values if local magnitudes are used instead of moment magnitudes.

Line 403: "probability of seismic hazard" -> "estimated seismic hazard"

Reply: We changed the sentence to your suggestion L456-L457: Higher b-values in the Raton Basin may indicate a decrease in the estimated seismic hazard, as they correspond to a lower likelihood of large earthquakes.

Lines 447-448: In CCT, ideally apparent stress values for calibration earthquakes should be constained using external information, which is now facilitated by the coda envelope ratio tool, CERT (a software companion to CCT). https://www.osti.gov/biblio/code-105734. Apparent stress is an individual earthquake property, not a regional property. Patton et al explored the sensitivity of Mw in cases where the apparent stresses of calibration earthquakes are unknown. Please clarify the text here.

Reply: Thanks for mentioning the CERT. We were not aware of such software. We added the following lines

- L499-L502:The CCT method assumes an apparent stress value, which is usually unknown for earthquakes in many regions, to constrain the high frequency part of the spectrum. Apparent stress values might be obtained with a companion software of the CCT, named Coda Envelope Ratio (CERT based on Walter et al. (2017)). Patton et al. (2025) show that, when the apparent stresses of the calibration earthquakes are unknown, selecting different assumed apparent stress values in the CCT methods affects the estimated MW for events with M < 3.

- We also added the website you provided in the Data and Resources Section.