Dear Editors of Seismica,

Thank you for taking the time to handle our manuscript during its revision process. We are happy to hear constructive and thorough responses from the reviewers. Generally, the critiques focused on clarifying our work and detailing other literature. We have endeavored to fully address these concerns through sentence revisions, wording corrections, and new citations that better outline the details of this work. Finally, we have also re-read the paper with fresh eyes to catch any lingering grammatical or typographic mistakes. This has prompted us to write slightly more about the results of the true-positive cases identified and the maximum magnitudes noted. We feel that these changes to the revised version adequately address the reviewer's comments, making the manuscript suitable for publication – We hope that you agree.

On the following pages are our itemized responses to the reviewer's comments. The original comments are in **bold-face***, my responses are in italics, and callouts to the original text are in "*blue-face.*"*

Thank you again,
- Ryan, Guillermo, Sebastián, & Javier

Reviewer #1:

I just finished revising the manuscript entitled "Chasing the ghost of fracking in the Vaca Muerta Formation: Induced seismicity in the Neuquén Basin, Argentina" by Schultz and co-authors submitted on Seismica.

First of all, I want to say that in my opinion the paper can be considered for a publication after addressing some comments. The paper is well written, discussed with several tests, also compared with other similar contexts. I cannot appreciate completely the pros and cons of the techniques without being an expert in this precise technique, but the overall judgment is positive from my side.

We'd like to thank the reviewer for their effort and detail that have been provided in reviewing our manuscript. We'd also like to take a moment to thank the reviewer for their overall positive outlook on this manuscript. Both are greatly appreciated!

The comments that I will show in detail in the two sections (major comments and minor comments) are basically, in my opinion, the basis to trust the analysis. In my opinion the Data section is a bit confusing, not properly explained. The authors give several information about the merge of the different catalogs they used, but at the same time the only information about the final dataset is the number of events. I would stress the quality of the dataset also because (my other comment) in paragraph 4.2 (limitation of the study) also the authors suggest an increase of the number of events to improve the results. But to appreciate the change one should have clear the starting point, so the quality of this dataset.

I also strongly suggest improving the conclusion section. It's a pity that all the discussion is summed up in a few weak sentences, which then defer to future work for a better understanding of what cannot be explained here.

We also appreciate the reviewer's thoughts here on how the manuscript could be improved. We have taken time to improve both the data and discussion sections. Partly motivated by the comments that were provided here, but also from our own re-read of the paper. Now we include more information on how the catalogues were concatenated and more on discussions. Below you can find a point-by-point response to each of the reviewer's more specific comments.

The following is the detail of the comments.

Major comments

Section 2.1 Earthqake catalogue

I found this section a bit confusing. I just propose small changes with the aim of make it clearer.

1) The authors mention as sources of data i) INPRES ii) a PhD study and iii) other studies. But then, the authors only explain the first two "sources". How have prior studies been taken into account?

This is a fair point for the reviewer to raise. The first two sources are the major ones. They constitute the bulk of the catalogue. We have slightly modified the wording here to be more accurate.

We note that all of these details can be found online at GitHub. There, we have a file called '20240314_VM_quakes_all_soruces.csv' under the 'data' subdirectory. This lists the source of the event put together in the INPRES+other catalogue. The other file 'final_base_phd.csv' gives the details of the PhD catalogue.

2) "The INPRES catalogue spans from 2015 onwards (predominantly starting in 2019), while the PhD project catalogue predominantly spans from 2014-2020."

a) What does it mean predominantly starting in 2019?

b) If it starts in 2015, the first catalogue does not have (or have a few) data between 2015 and 2019? If yes, why?

What this means: while there are some early events (2015-2018), this catalogue really starts seriously recognizing events by 2019. Largely because of station availability for catalogue building, between each of the catalogues (Figure S1). Said another way, the PhD catalogue has more/better data because it had access to additional stations during the early period (2014-2020). However, it is limited up until 2020. On the other hand, the INPRES catalogue spans the full duration, but only starts getting access to more local data from 2019 onwards. So, INPRES detects some events in 2015, but they're just far and few between.

To try and be more helpful to future readers, we've expanded the discussion of the catalogues. Specifically, a new paragraph has been added to Section 2.1 that covers the reviewer's comments here, as well as some of the other points raised throughout the review.

We'd also like to point out that the INPRES+other and PhD catalogues are available online, at GitHub for future readers to examine these points (or any others they may think of) themselves.

3) "The latter catalogue [...]. It comprised 11 stations covering an area of approximately 70×70 km 2 , which consistently collected data from November 2014 to July 2016. [...] From July 2016 until June 2020, the number of stations operational in the network was reduced to 5."

a) In the sentence before the authors state that the second catalog spans from 2014 to 2020 and then it seems that it has data only in the first part. Am I wrong? Otherwise I should assume that there are earthquake located with 5 stations? How good are these locations?

On this point we respectfully feel that the reviewer may have misunderstood something. It is correct to say the PhD catalogue spans 2014-2020: specifically, it uses 11 stations from Nov 2014 – July 2016 and then 5 stations from July 2016 – June 2020.

The events located with 5 stations are less ideal than those on 11. Recognizing this is part of the reason we've included a data limitations section to the paper.

4) "The limited overlap between catalogues means very few events are flagged as identical."

a) But from what I read, the period does not seem to me characterized by a limited overlap. Am I wrong?

b) If the authors made a choice, how did they do their choice? Based on a smaller error?

The reviewer is partly correct here. We have made some clarification that we think would be helpful.

Yes, these two catalogues are highly overlapped in time. However, they are not well-overlapping in their detections in time. The points mentioned above about station availability in building each of the catalogues (Figure S1) creates this disparity in detections.

*When a choice between 'identical' event(s) needs to be made, we have opted to use the PhD one, since it is generally more robust/accurate. We now mention this on Line 172 "*The preferred origin is selected as the one from the PhD project catalogue*".*

In conclusion, there are many details, but written in a bit confusing way, at least to me. If, in the author's opinion, the details on each contributor (INPRES, PhD etc.) to the catalogue is important, it shoul be better clarified the period, the number of events, the station configuration, and eventually some of the characteristics of the events (gap, rms, error ...). Then it would be better to declare which is the way the identical events have been selected. And finally highlight the main features of the final dataset: number of events, min-max magnitude, depth range, quality of the data. If a table can be helpful, the authors can also provide one. My comment is just for sake of clarity, that is crucial to trust the dataset.

To address this point, we have added a little more details to the paragraph at the end of Section 2.1. There we cover additional details on the distribution of earthquake depths. We do not mention minimum magnitude as this can be inferred from Figure 2.

Other more technical details like gap, rms error, # of observations, etc. are not reported in the maintext. Instead, readers can find these details in the plaintext file for the PhD catalogue (online at GitHub). Unfortunately, this information is not routinely reported/collected for the INPRES catalogue.

We would also like to highlight that we have included the code that was used to concatenate the two catalogues together online at GitHub as well. This way, any curious readers can recreate the catalogue (and the following results of the entire paper). They can also try different ways of combining the catalogues, if they'd like

Lines 433-443: it's nice to see in the "limitations of the study" a paragraph about the dataset (that links with my previous comments on the dataset issues).

Thanks! We felt this was appropriate given how sparsely recorded this region is for earthquakes.

a) For sure an improvement of the number and coverage of the seismic stations is crucial to reach smaller and smaller magnitude events, but it is crucial to understand why it should be improved by mentioning how many stations recorded the events and which is their quality (see comments on Dataset).

*Agreed. To illustrate this point better we have added the sentence "*In many cases, events in the Neuquén Basin were only recorded by a handful of stations, impacting the detection completeness and spatial resolution of the input catalogue*".*

b) The sentence at lines 436-440 seems just written to include in some way the machine learning in a generic way, but how can machine learning (which technique do you have in mind?) improve the dataset? There are example of template-matching improving the dataset? How it works? I suggest deepening the sentence that is quite general.

*Fair point, we have now clarified that we were referring to the use of machine learning during the phase picking process. We have also added another sentence (Lines 461-462) that says, "*Both techniques have shown significant promise for improving catalogue sizes by up to an order-of-magnitude*". To give the reader a bit more context as to what is meant here.*

c) "Overall, improving dataset quality will be important for the future of earthquake risk management in the Neuquén Basin.". I agree, but make it more quantitative, starting from how is it the quality of this dataset (see comments on dataset)

On this point, we have opted to cite a new review paper [Zhou et al., 2024] that details how risks from induced seismicity can be best managed. This paper goes into full detail as to how different topics are interrelated, towards the common goal of risk management – one topic of which is earthquake monitoring and catalogue building.

Zhou, W., Lanza, F., Grigoratos, I., Schultz, R., Cousse, J., Trutnevyte, E., ... & Wiemer, S. (2024). Managing induced seismicity risks from enhanced geothermal systems: A good practice guideline. Reviews of Geophysics, 62(4), e2024RG000849, https://doi.org/10.1029/2024RG000849.

Conclusion

My personal opinion is that the conclusions are weak. The analysis is strong, there are tens of tests, there are numbers coming out from the analysis, supporting the inference and the discussion is quite interesting. I agree that the analysis probably (I am not an expert of the technique) is done on a few events, from the statistical point of view, and the dataset and the quality of the dataset should be improved, that is always a good idea. But there are no numbers and no results in the conclusion, there is not a usual summary of what has been done and it ends with "This lack of clarity could be

resolved with additional data (and then further study).", in my opinion not the best way to close a paper. I suggest to be more quantitative, to show what has been done and propose of course an improvement, but based on what you achieved here.

We thank the reviewer for pointing this out. We have revised the conclusions paragraph to include more of the quantitative details of the paper. We have also added a new sentence to the conclusions to provide richer details on the true-positive cases identified.

Minor comments

Line 67: the sentence starts with "for example" but it refers to the previous sentence. I would choose. I would either link the two sentences together or start this one with, for example "Already many examples are in the literature, such as ...".

In this particular case we have opted not to make this change. The prior sentence refers to a general statement. Then this sentence (Line 69), gives a list of examples of that general statement. Because of this, we feel that the use of "for example" is appropriate here.

Line 161: cotemporaneous with contemporaneous

Corrected. Thanks to the reviewer for catching this.

Line 173: should you add the website where you downloaded the data, if it is free?

To this point, we respectfully point the reviewer to the "Open Research" section of the paper. There, we provide the hyperlinks for the INPRES earthquake catalogue and the Argentina wells. In addition to this, we have also provided files of each dataset used in this study, ready for download online at GitHub.

Line 392: On the PDF it appears " 'YPF.Nq.LLLO.x-2(h)' ". Be careful, I don't know if it is my problem or from the source text.

We thank the reviewer for pointing this out. Note that this is correct. The unique well identifiers for Argentina follows this format. An example of this can also be seen alongside the well callouts on Line 301. Note that this also corresponds to the 'SIGLA' field in the hydraulic fracturing databases (that's available online).

Reviewer #2:

The author explained the sources of the datasets, described the method used to screen for induced events (a spatiotemporal association filter), and analyzed the robustness of the results. The author also quantified the strength of the time-lagged correlation between seismicity and hydraulic fracturing injection rates. Additionally, the author discussed comparisons with known cases caused by hydraulic fracturing and pointed out the limitations of this research.

This manuscript is complete, well-written, and well-organized. However, I would like to provide the following comments before it can be accepted:

We'd like to thank the reviewer for their time and effort taken in critiquing our manuscript. We are also grateful for the kind words that have been provided here. See our comment below for the response to the reviewer.

Please expand the review section to include literature on other methods for identifying induced earthquakes, particularly those related to the ETAS method [1,2]. Additionally, please provide a brief analysis of the differences and connections between these methods concerning the identification of induced earthquakes.
I hope these additions will strengthen the manuscript and provide valuable insights for readers.

[1] Liu, Y., Zhuang, J., Guo, Y., Jiang, C., Tian, Q., Zhang, Y. and Long, F., 2024. Background and clustering characteristics of recent seismicity in Southwestern China. Geophysical Journal International, 238(3), pp.1291-1313.
[2] Petrillo, G., Kumazawa, T., Napolitano, F., Capuano, P. and Zhuang, J., 2024. Fluids-Triggered Swarm Sequence Supported by a Nonstationary Epidemic-Like Description of Seismicity. Seismological Research Letters, doi: https://doi.org/10.1785/0220240056.

*Certainly. I have added a citation to the SRL paper in the discussion section now, at the point where "*advanced methods*" for discriminating induced seismicity is mentioned.*

Dear Editor

Dear Editor,
I just finished revising the manuscript entitled “Chasing the ghost of fracking in the Vaca Muerta Formation: Induced seismicity in the Neuquén Basin, Argentina” by Schultz and co-authors, reviewed and resubmitted on Seismica.

My general comment is still positive, and the manuscript for sure deserves to be published on Seismica. I leave some more very small comments, also based on the answers to my previous comments. Most of the comments have been addressed but, from my personal point of view, the conclusions are still a bit weak, and I would fix and better clarify the quality of the dataset.

Data: Probably, for the latter, I am too picky and maybe is not needed for the statistical analysis, but trusting the dataset is crucial, from my side. So I leave to the authors (and of course to the Editor) the choice of improving this part.

Conclusions: The conclusions need to be strengthened, the sentences supported by the results, not just a few numbers. The last sentence is way too weak and probably not needed.

Detailed comments

Line 160: change “catalouge” with “catalogue”

Lines 160-165: too many times there is the repetition of the word “catalogue”. Please be careful of the style here.

Line 168: “The preferred origin is selected as the one from the PhD project catalogue.“ This has been added following my previous comment, but it is a qualitative statement that does not strengthen the discussion. I do not want to be so picky on the catalog quality, but it is crucial to rely on the results.

Line 448-450: here again qualitative statement “a handful of stations”. Numbers are important to trust the results, especially in locating earthquakes. Locations with very few stations cannot be trusted, mainly at this scale, that is at fracture/small magnitude event scale.

Lines 450-454: I still think, as in my previous comments, that this sentence as it is does not add anything to the paper. The authors did not apply these techniques, just mention them. Either they better constrain these sentences, supported by evidence (as for example supporting with references, as Petrillo et al. 2024, suggested by the reviewer 2) or remove it.

Line 455: Petrillo et al. 2014 is actually Petrillo et al., 2024

Lines 460-470: I still think that the conclusions are weak. Following my previous comments the authors added some numbers, but this does not strengthen the conclusions, that are way qualitative. The technique is not mentioned, not the novelty of this paper. There still are some weak sentences, as also in the case of the very last sentence.