July 16, 2025

Editor-in-Chief

Natural Hazards and Earth System Sciences

Dear Editor,

I am pleased to submit this cover letter regarding the original research article (nhess-2024-102) entitled “Could seismo-volcanic catalogues be improved or created using weakly supervised approaches with pre-trained systems?” by Titos M., et al., for consideration in NHESS. We have carefully reviewed the feedback from **all two reviewers** and greatly appreciate the time and effort they have invested in evaluating our work.

We hope that these minor revisions, alongside the provided documentation of changes, meet the reviewers’ expectations and adequately address their feedback.

Thank you for the opportunity to improve our work.

Yours sincerely

Manuel Marcelino Titos Luzón
Postdoctoral Researcher, University of Granada, Spain
mmtitos@ugr.es.

ANSWER TO THE REVIEWER'S COMMENTS

In the following, we have provided detailed answers to the comments of the reviewers. The original texts from the reviewers are in normal font. Our answers are in bold font. We would like to take this opportunity to thank the reviewers for their valuable comments and for their time and resources.

Answer to comments of Reviewer#1

We would like to thank reviewer#1 for the careful reading of this manuscript and the thoughtful comments that have improved the quality of this manuscript. Furthermore, below are those comments that need more clarification.

Please comment on how volcano-specific the results are, and how relevant they are to other volcanoes.

To clarify this important aspect of the methodology and the method's applicability, we have modified Section 5.4 of the manuscript. In the revised version, we address the specificity and relevance of the master database in shaping the final results. We have included the following explanation: "One of the main strengths of the proposed system is its ability to recognize previously learned prototype events, even in volcanic environments that differ significantly from those present in the training datasets. This feature enhances its usefulness in reducing biases when creating or updating catalogs. The results suggest that training on a broader variety of volcanic settings with diverse event prototype distributions could improve recognition performance, fostering the development of more generalizable and less biased catalogs. Nonetheless, the system also presents limitations. Since the pseudo-catalogs are generated using models trained on a fixed set of known seismic categories, the system is forced to assign one of these categories to each analyzed window, even when the event does not match any known prototype. This constraint can lead to the mislabeling of truly novel events and, consequently, affect the performance of systems retrained using such pseudo-labels. Addressing this issue would require the creation of more comprehensive master databases that incorporate a wider range of event types, ideally from multiple volcanic settings. Moreover, determining the appropriate membership threshold for including events in the new pseudo-catalogs remains a key challenge. Low thresholds may increase sensitivity but also introduce many false positives—events that are dissimilar to any known prototype. Retraining the systems with these catalogs could reduce performance and detection accuracy. High thresholds, on the other hand, may improve specificity but may not be sufficient to allow the system to adapt to the new volcanic environment. This trade-off highlights the importance of post-analysis tools that assess detection confidence, which, in

addition to offering insights into the presence of potentially novel classes not covered by the original training data, also contribute to evaluating the reliability and effectiveness of the domain adaptation process by revealing how well the system distinguishes between learned and unfamiliar patterns in new volcanic environments—that is, how volcano-specific the results are and how relevant they may be to other volcanoes."

Answer to comments of Reviewer#2

Dear reviewer#2, we sincerely appreciate your thorough evaluation of and the valuable suggestions, which have significantly contributed to enhancing the clarity and quality of the manuscript.

The authors have addressed the major concerns raised in previous review rounds, providing clarifications, methodological adjustments, and additional analysis that substantially improve the clarity and scientific value of the manuscript. The experiments are now better contextualized, the performance metrics have been calculated as requested, and the discussion reflects a more balanced interpretation of the model's strengths and limitations. With these revisions, I believe the manuscript is suitable for publication and will be a valuable contribution to the field of machine learning in seismology.

As a final recommendation, I suggest that the authors consider including the precision, recall, and F1-score metrics—either in the appendix or as supplementary material—to enhance transparency and allow for easier comparison with related studies.

In this revised version of the manuscript, the authors have included F1- score metrics in the supplementary material.