RC1: 'Comment on egusphere-2026-3012', Anonymous Referee #1, 09 Jul 2026

Revision of “Detecting Fault Structures from Earthquake Sequences via Unsupervised Learning”

Comments to the Author(s):

Comments 1: The manuscript presents a pipeline that combines DBSCAN unsupervised clustering and PCA analysis to identify potential fault structures from the 2018 and 2024 earthquake sequences in Taiwan. The topic is relevant, and the approach could be useful for analysing complex aftershock distributions, but the manuscript requires major revision before publication. The main weakness is that the proposed method is not sufficiently justified, validated, or tested. The authors should clarify some aspect such as how DBSCAN and PCA are combined, whether PCA is applied separately to each DBSCAN cluster, how the PCA-derived fault planes are defined, and how sensitive the results are to the selected DBSCAN parameters. This is particularly important because, in at least one case, the parameter set with the highest Silhouette Score was not adopted, and a more geologically plausible solution was chosen instead. This choice may be reasonable, but it reduces the objectivity of the workflow and requires a sensitivity analysis.

Response 1: We appreciate the constructive comments and will incorporate a flowchart to illustrate the applications of DBSCAN and PCA, thereby clarifying their respective roles at different analytical stages. The Silhouette Score (SS) serves as the primary clustering validity index, accounting for both ε and *MinPts* parameters. Selecting the highest SS value generally yields the optimal clustering result; however, the number of clusters must also be considered to ensure the delineation of potential fault planes.

For instance, in the EQ2018 case (Figure 3a), the maximum SS value (ε = 0.08, *MinPts* = 11) produces only two clusters, grouping both inland and offshore seismic events together within the same cluster. By contrast, the second-highest SS value (ε =

0.07, *MinPts* = 10) increases the number of clusters to four, thereby enhancing the identification of potential fault planes.

In addition to SS, we will incorporate supplementary indices for comparative evaluation, including Density-Based Clustering Validation (DBCV), Davies–Bouldin Index (DBI), and Calinski–Harabasz Index (CHI). These results will be integrated into the revised manuscript to strengthen the robustness of the clustering assessment.

The clustering of seismic events inevitably influences the subsequent PCA computations, as different spatial distributions of earthquakes yield distinct fault-plane orientations. We will further elaborate on these results by clarifying how the completed focal mechanism analyses corroborate the PCA-derived structures. Specifically, the integration of clustering outcomes with PCA, supported by the focal mechanism interpretations, enhances the delineation of latent fault geometries and strengthens the reliability of the identified planes. This extended discussion will be incorporated into the revised manuscript.

Comments 2: Another issue concerns also the use of PCA-derived fault-plane areas to estimate potential maximum magnitude. This step is not sufficiently motivated, and it is unclear why it is necessary for the clustering analysis. A PCA plane is a statistical approximation of an aftershock cloud, not necessarily a real rupture surface. Therefore, small variations in the inferred plane dimensions may strongly affect the estimated magnitude and could lead to misleading implications for seismic hazard. Also the location error could influence the PCA analysis results. The authors should either validate this approach using well-constrained earthquake sequences, provide uncertainty estimates, and cite previous studies using a similar method, or remove/reframe this part as a preliminary and exploratory analysis. The magnitude–area relationship should also be checked and clearly described.

Response 2: The primary motivation for estimating the potential maximum magnitude is to identify structural weaknesses that may serve as loci for future rupture. Fault rupture can be conceptualized as an earthquake event, occurring at zones of structural weakness beneath a specific region (King et al., 1994; Stein, 1999; Freed, 2005; Toda et al., 2005).

The clustering results in this study reveal plane-like distributions of seismicity, some of which align with known tectonic structures, while others diverge from established geological frameworks. We acknowledge that the quality of earthquake location data is a critical factor in accurately delineating fault planes. To enhance clarity, the revised manuscript will provide further details on the earthquake catalogue and the magnitude–area scaling relationship employed in this analysis.

Moreover, by integrating clustering-derived geometries with magnitude estimation, this framework not only highlights potential structural weaknesses but also underscores the importance of data precision in constraining seismogenic potential. Such an approach strengthens the interpretive link between observed seismic distributions and the identification of latent fault structures, thereby contributing to both scientific understanding and hazard assessment.

Comments 3: The comparison with known fault systems also needs to be more rigorous. At present, it is mostly descriptive. An additional figure showing mapped faults, PCA-derived planes, clusters, and focal mechanisms would strengthen the validation of the method proposed. Also the visualization of only one focal mechanism per cluster is not sufficient to validate the inferred fault geometry; more focal mechanisms or additional independent constraints should be included.

Response 3: We appreciate the valuable comment. In response, we will include an additional figure that integrates mapped faults, PCA-derived planes, clustering results, and focal mechanism solutions. Wherever possible, focal mechanism solutions will be displayed for each cluster, accompanied by explanatory text to ensure consistency between graphical and textual presentation.

This enhancement will provide a more comprehensive visualization of the structural framework, allowing readers to directly compare clustering-derived geometries with both mapped tectonic features and focal mechanism interpretations. By presenting these elements together, the revised manuscript will strengthen the interpretive link between statistical clustering, PCA-derived planes, and seismological observations, thereby improving the clarity and robustness of the analysis.

Comments 4: The manuscript should also discuss uncertainties more explicitly, including earthquake-location errors (if available), uncertainty in PCA-derived strike and dip, uncertainty in plane dimensions, and the effect of using Min-Max normalization on the geometry of the inferred planes. Since longitude, latitude, and depth have different physical scales and uncertainties, normalization may influence clustering, PCA orientation, and the resulting magnitude estimation.

Response 4: We acknowledge that the quality of earthquake location data is a critical factor in accurately delineating fault planes. To enhance clarity, the revised manuscript will provide additional details on the earthquake catalogue. In PCA, the selection of principal components is based on the variance derived from the singular value decomposition (SVD) solution, which is constructed according to earthquake locations. The leading three components (PC1, PC2, and PC3) are considered in fault-plane estimation. The computational procedure for determining the best-fit plane will be updated in the revised version to ensure methodological transparency.

Furthermore, Min–Max normalization is applied to mitigate the influence of variables with differing physical scales prior to data processing, thereby ensuring balanced weighting across dimensions. This normalization step, combined with PCA, provides a robust framework for translating discrete hypocentral distributions into geologically meaningful fault-plane geometries. The revised manuscript will expand on this workflow to highlight how data quality, normalization, and component selection collectively strengthen the reliability of structural interpretations.

Comments 5: The interpretation of temporal migration in the 2024 sequence should also be better supported. The distinction between T1 and T2 needs to be clearly defined in the text and figures, and any interpretation in terms of stress redistribution or secondary fault activation should be supported by references and more quantitative analysis. Finally, the manuscript needs clearer writing, more references in the introduction and seismicity description, improved figure captions, and better explanation of acronyms and plotted quantities. Several figures should be improved in readability, with larger fonts, clearer labels, visible acronyms, and a better representation of outlier/noise points.

Response 5: We appreciate the reviewer's insightful suggestions regarding the interpretation of temporal migration in the 2024 sequence. In the revised manuscript, we will strengthen this section by:

- Clearer definition of T1 and T2: The temporal windows will be explicitly defined in both the text and figures, with improved captions to ensure clarity.
- Support for interpretation: We will expand the discussion of temporal migration by incorporating references to established studies on stress redistribution and secondary fault activation.
- Improved manuscript presentation: The introduction and seismicity description will be supplemented with additional references to strengthen the context. Acronyms and plotted quantities will be explained more thoroughly in the text and figure captions.
- Figure readability: Several figures will be revised with larger fonts, clearer labels, visible acronyms, and improved representation of outlier/noise points to enhance readability and interpretability.

Comments 6: Overall, the current version does not yet demonstrate that the DBSCAN–PCA workflow is robust enough to infer fault geometries and potential magnitudes. I therefore recommend major revision, with substantial methodological clarification and validation required before publication.

Response 6: We sincerely appreciate the reviewer's careful evaluation and acknowledge the concern that the current version does not yet fully demonstrate the robustness of the DBSCAN–PCA workflow for inferring fault geometries and potential magnitudes. In the revised manuscript, we will undertake the following substantial improvements:

- Methodological clarification: We will expand the description of the DBSCAN–PCA workflow, including parameter selection, normalization procedures, and computational steps for plane estimation, to ensure transparency and reproducibility.
- Validation and robustness: Additional tests will be incorporated to validate the workflow, including comparisons with mapped faults, focal mechanism

solutions, and alternative clustering indices (e.g., DBCV, DBI, CHI). These will demonstrate the consistency and limitations of the approach.

- Magnitude estimation: The rationale for estimating potential maximum magnitudes will be clarified, with explicit reference to the earthquake catalogue and magnitude–area scaling relationships. This will strengthen the link between clustering results and seismogenic potential.
- Presentation and clarity: Figures will be revised for readability (larger fonts, clearer labels, acronyms, improved representation of outliers/noise points), and captions will be expanded to provide sufficient context. Acronyms and plotted quantities will be explained more thoroughly in the text.

A few questions/comments/suggestions:

Comments 7: Lines 7-8: You start you abstract with the acronyms of the earthquakes. I suggest writing information about it and then refers it using the acronyms. EQ2018 and 2024 are cited the first time at lines 102.

Response 7: We will update the words.

Comments 8: Lines 49-50: I think that indicated date and magnitude of the mainshocks could help the reader in the understanding.

Response 8: We will update the words.

Comments 9: Lines 102-130: Please insert references when you describe the two seismic sequences.

Response 9: We will update the words.

Comments 10: Lines 144-146: Cite who create these algorithms.

Response 10: We will update the words.

Comments 11: Lines 209-212: You are using PCA to determine the length of the fault plane. A small variation in length can affect the expected Mw estimate. For this reason, the length must be as accurate as possible in order to avoid problems and unnecessary alarm. Are there any studies that have used this innovative approach? If so, I recommend citing them to strengthen your work; otherwise, I would first test the approach on a sequence for which I know the Mw values well and that has been thoroughly verified, to ensure the approach works before proposing it as a solution. Also, what values are being used—PC1, PC2, or PC3? Please specify this in the text. You should also specify why you compute the potential magnitude. What is its purpose in the seismicity clustering phase? Is the clustering spatial, or is there also clustering based on magnitude? Please specify.

Response 11: In this study, the fault is idealized as a rectangular plane, whose orientation is determined using the leading principal components (PC1, PC2, and PC3). The spatial length and width of the plane are estimated from the distribution of clustered seismicity, effectively representing the best-fit plane to the earthquake locations. Wells and Coppersmith (1994) established empirical scaling laws linking earthquake magnitude to fault parameters such as length, width, and area. In this work, we adopt the magnitude–area scaling relationship to infer the maximum potential magnitude from the estimated fault area.

To strengthen methodological transparency, the revised manuscript will provide additional details on the computational procedure for plane fitting and the assumptions underlying the rectangular fault representation. This clarification will highlight how clustering-derived geometries are systematically translated into quantitative estimates of seismogenic potential, thereby reinforcing the robustness of the DBSCAN–PCA workflow.

Comments 12: Lines 264-265: Interval? Is the T1 and T2? It's difficult to understand if you cite in one way in the figure and then in a different way in the text….

Response 12: We will update the words.

Comments 13: Lines 362-396: I would have expected a more rigorous comparison with the faults documented in the literature, rather than just a descriptive one (perhaps a new figure showing the overlay of the faults from the literature with those modeled using PCA). This paragraph does not seem to do justice to the work that was done, and a summary image showing the comparison would also be useful to support the procedure. Especially since the clustering changes when the epsilon and minpnt values are modified. There is also no geological interpretation of the faults modeled using these techniques...

Response 13: We will update the words. Also, we will include an additional figure that integrates mapped faults, PCA-derived planes, clustering results, and focal mechanism solutions. Wherever possible, focal mechanism solutions will be displayed for each cluster, accompanied by explanatory text to ensure consistency between graphical and textual presentation.

Comments 14: Lines 404-406: The geometry of a fault is determined not by a single focal mechanism (as shown in Figure 6) but by many more. Either you create a diagram that includes many focal mechanisms to confirm the model you've developed, or—put that way—it does not serve as evidence that the work was successful.

Response 14: We will incorporate additional focal mechanism solutions for each cluster, accompanied by explanatory text to ensure consistency between graphical and textual presentation. This integration will allow readers to directly compare clustering results with focal mechanism interpretations, thereby reinforcing the robustness of the structural analysis.

Comments on Figures:

Comments 15: Figure 1: Increase the font and the dimension of the map on the right. I suggest also insert the acronyms in the Figure. I don't understand the T1 and T2 meaning for 2024. Is it described in the text? You mean that in the EQ2024 the seismicity moved away from the mainshocks? You should explain the eventually migration in the text.

Response 15: We will update Figure 1 and relevant words.

Comments 16: Figure 2: If I understood well, you divided the period T1 and T2 observing Fig. 1b. If yes, create two windows (maybe square shape) in the 1b and write “T1” and “T2”.

Response 16: We will update Figure 2 and relevant words.

Comments 17: Figure 3: Why there are many number on the x axes? You can also show the outlier points in the map. Specify the acronymous name (Dstd, Dmean…)

Response 17: We will update Figure 3.

Comments 18: Figure 4: Interval? Is T1 and T2? Specify.

Response 16: We will update words for Figure 4.

Comments 19: Figure 6: I suggest to specify the magnitude of the focal mechanism in the figure and in the text.

Response 19: We will update Figure 6 and relevant words.