

Reviewer #2

To allow a better communication of our replies, we have divided them into three categories: Agree/Clarification/Disagree.

1.

Suggestion, Question, or Comment from the Reviewer #2	Author's Response	Change in the Manuscript
This paper tests three different methods of scaling input data to a machine learning workflow. The methods are tested for their ability to maintain spatial additivity, meaning to produce the same results when applied on a subregion of a spatial dataset as when applied to the entire dataset. The authors conclude that canonical machine learning scaling, in which scale factors are fit to the training data and then applied to all datasets (training, synthetic test, and real-world data), is the only one of the three methods that meets the criterion of spatial additivity. The scaling methods are tested on a machine learning workflow for fault detection from one of the second author's previous papers, but the conclusions are likely relevant to applications of machine learning to spatial data more broadly. The methods and reasoning are sound as far as I can tell, and I found the paper mostly clear, but I do think some minor to	We thank Reviewer 2 for their report and spotting the clear mistakes and unclear statements. It encouraged us to perform a more comprehensive revision of language.	Not applicable.

moderate revision is required, as detailed in my comments below.		
--	--	--

2.

Suggestion, Question, or Comment from the Reviewer #2	Author's Response	Change in the Manuscript
It would be good to have an explanation of how the three scaling methods were chosen. Especially since the “canonical” method turns out to be the best one, it is worth discussing why other methods were being considered to begin with. In the background section, several other papers are cited that have used machine learning for geological feature detection. Perhaps stating how these papers have scaled data (if known) would help to justify the need for a comparison of different methods.	Agree/Clarification We agree with this comment. 1. The first method (Method 1) was used by Michalak et al. (2025) - see the reference below. 2. The second method is motivated by the research into application of ML in biomedical engineering. We quote a paragraph from Apicella et al. 2023 (see the references below): <i>In Chen et al. (2021b) several normalization schemes were applied following two different schemas:</i> <i>(i) All-subjects, where the whole dataset was normalized,</i> <i>(ii) Single subject, where the normalization is made individually for each subject.</i> 3. The “canonical ML scaling” adheres fully to ML standards, so it can be considered a “textbook” approach.	We provided a clarifying reference in Introduction. We've added a reference Chen et al. 2023 which differentiates between all-subject and single subject approaches. We added a clarification to section 6.1 about similarities and differences with the existing normalization approaches. For clarity, we removed some synonymous terms for “all-subject” and “single-subject” schema.

--	--	--

Reference:

1. Michalak, M., Gerhards, C., Menzel, P., (2025). SubsurfaceBreaks v. 1.0: A supervised detection of fault-related structures on triangulated models of subsurface homoclinal interfaces. Geoscientific Model Development, 18, 4469–4481, <https://doi.org/10.5194/gmd-18-4469-2025>
2. Apicella, A., Isgrò, F., Pollastro, A., and Prevete, R.: On the effects of data normalization for domain adaptation on EEG data, Eng. Appl. Artif. Intell., 123, <https://doi.org/10.1016/j.engappai.2023.106205>, 2023.
3. Chen, H., Sun, S., Li, J., Yu, R., Li, N., Li, X., Hu, B., 2021b. Personal-zscore: Eliminating individual difference for EEG-based cross-subject emotion recognition. IEEE Trans. Affect. Comput..

3.

Suggestion, Question, or Comment from the Reviewer #2	Author's Response	Change in the Manuscript
While “spatial additivity” is defined early in the paper, a reference to or discussion of previous work on the concept would be helpful.	Agree/Clarification We found some discussions about spatial additivity for ecological studies. For example, in one of the studies, spatial additivity meant that the diversity of a studied area (e.g. region) is the sum of the local diversities. We've also found research which addresses the problem of predicting beyond the geographic boundaries of training data.	We've referenced the definition of spatial additivity from ecological studies (Granger et al. 2015). We've added a discussion from Meyer&Pebesma 2021 (see the reference below) in section 3.6.

References:

Meyer, H., & Pebesma, E. (2021). Predicting into unknown space? Estimating the area of applicability of spatial prediction models. *Methods in Ecology and Evolution*, 12(9), 1620-1633.

Granger, V., Bez, N., Fromentin, J. M., Meynard, C., Jadaud, A., & Merigot, B. (2015). Mapping diversity indices: not a trivial issue. *Methods in Ecology and Evolution*, 6(6), 688-696.

4.

Suggestion, Question, or Comment from the Reviewer #2	Author's Response	Change in the Manuscript
The use of spatial additivity as an evaluation criterion seems to me to be one of the most broadly transportable ideas from this paper. Applications of machine learning in geology almost always deal with spatial data, and I have not seen others test for spatial additivity. Some additional discussion of this criterion and its value beyond the present study would strengthen the paper.	Agree/Clarification We thank the reviewer for this constructive comment. Our current implementation evaluates spatial additivity within the context of classification (e.g., fault mapping). However, the core underlying concept can be extended to spatial regression problems, such as predicting continuous parameters like depth-to-bedrock (Pawley et al. 2024 - see the below reference). In regression, spatial non-additivity would manifest as numerical step-discontinuities across boundaries of the subsets rather than label changes. We consider this extension as an important avenue for future research.	We've added a paragraph about possible extension to regression problems in Section 6.6 (just before section 6.7).

References:

Pawley, S. M., Atkinson, L., Utting, D. J., Hartman, G. M. D., and Atkinson, N.: Evaluating spatially enabled machine learning approaches to depth to bedrock mapping, Alberta, Canada, PLoS One, 19, <https://doi.org/10.1371/journal.pone.0296881>, 2024.

5.

Suggestion, Question, or Comment from the Reviewer #2	Author's Response	Change in the Manuscript
Lines 144-146: This reasoning isn't clear to me. Wouldn't the solution to different ranges of different attributes be to scale each data type independently, rather than each data set (which presumably includes data of all types)? Indeed, doing that seems to have already been discussed in the Introduction.	Clarification We agree that we made redundancy (different data ranges already discussed in Introduction) and confusion (features and not the entire data sets were scaled).	We have rewritten the corresponding paragraph (section 3.2- Method 1: Independent subset scaling).

6.

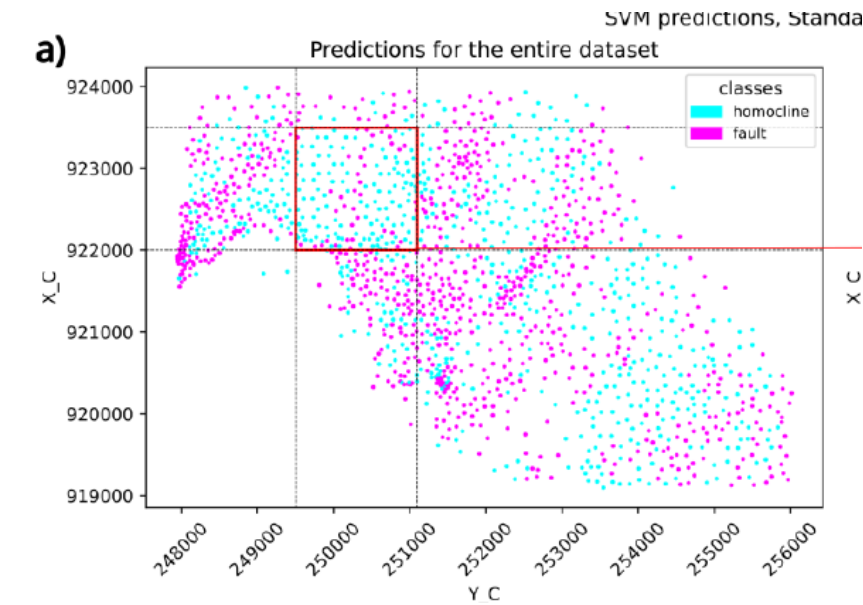
Suggestion, Question, or Comment from the Reviewer #2	Author's Response	Change in the Manuscript

Line 374-5: Is it possible to say which known faults these were? Are they ones in the maps or cross sections shown earlier?	Clarification We agree that the information about “known fault” was missing. We meant the potential N-S trending fault visible in Fig. 8 a in the red rectangle (see the below picture, as well) and already discussed in Michalak et al. 2025 (Fig. 7a there), and Michalak et al. 2022 (Fig. 11b, there).	We have provided more information about the “known faults” in section 6.2 - Spatial additivity and interpretability of predictions, spatial autocorrelation ”.
---	---	---

References:

Michalak, M. P., Gerhards, C., and Menzel, P.: SubsurfaceBreaks v. 1.0: a supervised detection of fault-related structures on triangulated models of subsurface homoclinal interfaces, Geosci. Model Dev., 18, 4469–4481, <https://doi.org/10.5194/gmd-18-4469-2025>, 2025.

Michalak, M. P., Teper, L., Wellmann, F., Zaba, J., Gaidzik, K., Kostur, M., Maystrenko, Y. P., and Leonowicz, P.: Clustering has a meaning: optimization of angular similarity to detect 3D geometric anomalies in geological terrains, Solid Earth, 13, 1697–1720, <https://doi.org/10.5194/se-13-1697-2022>, 2022.



Suggestion, Question, or Comment from the Reviewer #2	Author's Response	Change in the Manuscript
Line 454: It's unclear to me what the "local re-triangulation" refers to. Aren't the models already triangulated as input to the ML algorithm, as shown in Fig. 1?	Clarification In this paragraph, we described updating geological models due to new data. So when new data should be incorporated in the model, updating the triangulation model is necessary (a new triangulation is needed).	We have rewritten the sentence for better clarity.

8.

Suggestion, Question, or Comment from the Reviewer #2	Author's Response	Change in the Manuscript
Minor Comments: Line 57: Should "test and production tests" be "test and production sets?"	Agree We thank the Reviewer for spotting the typo.	We've corrected the typo.

9.

Suggestion, Question, or Comment from the Reviewer #2	Author's Response	Change in the Manuscript

Line 122: the word “the” is missing before “following.”	Agree We agree that the word “the” was missing.	We’ve corrected this mistake in section 2.2.
---	---	--

10.

Suggestion, Question, or Comment from the Reviewer #2	Author’s Response	Change in the Manuscript
The heading to section 5.2 should not end in “and.”	Agree We agree: “and” was redundant.	We’ve deleted “and” (it is now section 5.3)

11.

Suggestion, Question, or Comment from the Reviewer #2	Author’s Response	Change in the Manuscript
For part c of the Fig. 9 caption, I think this should say “red rectangle” not “red triangle.”	Agree We agree about the mistake.	We’ve replaced “triangle” with “rectangle” in the caption of Fig. 9.

Our own changes:

Section 6.6 (Bridging the research gap) - We’ve removed some strict claims from Discussion such as “lack of discussion”. Instead, we propose to replace it with “limited

discussion”, since the Reviewer #1 provided a context (spatial autocorrelation) which allowed us to find more papers dealing with leakage but mostly in soil and hydrological science.

Nevertheless, we believe that it may be true that the discussion about leakage is very limited in seismic ML modeling because in the referenced review paper the term leakage has not been mentioned: “A systematic review of deep learning for structural geological interpretation”.

Fernandes, G. L., Figueiredo, F., Hatushika, R. S., Leao, M. L., Mariano, B. A., Monteiro, B. A. A., ... & Zavam, F. (2025). A systematic review of deep learning for structural geological interpretation: GL Fernandes et al. *Data Mining and Knowledge Discovery*, 39(1), 3.

We’ve corrected some typos and mistakes about best hyperparameters/numbers in tables (A1-A6). We’ve also performed a more comprehensive revision of language using ChatGPT (grammar, redundant expressions, awkward phrasing). The monitoring of all changes is possible using the marked version of the manuscript.