



EGUsphere, referee comment RC1
<https://doi.org/10.5194/egusphere-2022-1190-RC1>, 2023
© Author(s) 2023. This work is distributed under
the Creative Commons Attribution 4.0 License.

Comment on egusphere-2022-1190

Lukas Mosser (Referee)

Referee comment on "Complex fault system revealed by 3-D seismic reflection data with deep learning and fault network analysis" by Thilo Wrona et al., EGU Sphere,
<https://doi.org/10.5194/egusphere-2022-1190-RC1>, 2023

Dear Authors, dear Editor,

Thank you for the opportunity to review this manuscript.

I have read it with great interest, and find the manuscript well written, comprehensive, and an important study showing the potential of machine learning-enabled large-scale studies of geological structures in application to CO₂ sequestration. While the proposed study is centred around subsurface storage of CO₂, I am certain that similar methodologies could be applied in other areas e.g. in near-surface geophysics.

I will first outline a brief summary of the main findings, and then provide a few high-level questions/comments which I hope will further improve the manuscript. For this, I will mainly focus on the machine learning aspect, and fault extraction aspect of the presented work, as this is where I see the area that I am most confident in my own understanding. Nevertheless, I may comment on the geological aspect, and highlight here that my comments are well-intended, but misinformed due to a lack of expertise in a field in which I am confident the authors and other reviewers are well-versed.

The presented manuscript proposes the application of deep neural networks and automated fault extraction to perform rapid large-scale analysis of normal faults from 3D seismic data. The study area is focused on the North Sea covering roughly 35000 km² of the northern North Sea rift zone. Deep neural networks have been trained on 2D patches extracted. The authors use an automated fault extraction method to ultimately analyse fault length, strike, density, and continuity. In their discussion, they highlight the advantages of deep learning-based fault interpretation and also relate this to manual, human interpretation. The authors show comprehensively how these technologies can in combination generate new insights into geological systems that would otherwise be extremely tedious to perform by manual interpretation while providing a repeatable

process that can be automated.

The manuscript is accompanied by a number of well-laid-out figures that support the findings described in the text.

In summary, I do not see any major issues with the manuscript and it should therefore be accepted subject to minor revisions and comments.

Line 29: Would the same study be possible for other types of faults? Have only normal faults been labelled in the data labelling process for training the deep neural network? Would the network be able to distinguish normal faults from reverse faults for example?

Line 46 and 47: I feel the wording around "proof of concept" and "has yet to" could be improved. There is a lot of groundwork that needs to be done to understand the ability of these deep neural networks to become a reliable source of knowledge. Just because there has not been a study to evaluate the insights gained does not mean there wasn't potential to do so coming from numerous works where the ability of the networks to detect faults have been established, some of which have even been published including Model weights e.g. Wu et. al.

Line 50: I am not sure I like the use of "<0.1% of data volume". I have made this statement myself before, but I believe rather than focusing on the reduced volume we should focus on data quality. Apart from the fact that seismic data have strong lateral correlations making additional neighbouring data less diverse, we can also consider other criteria: How many of the relevant types of faults have been mapped, and how many noise modalities have been incorporated? Nothing to change here for now, but potentially something to address in the future?

Line 88: Accuracy is a metric that is not well suited for class imbalanced problems such as fault detection. Have you considered using the F1 score?

Line 89: Did you monitor the training or validation loss?

Line 93: Have you considered also publishing the weights?

Line 108: The faults identified at threshold < 0.3 may be small faults but also misclassifications, the same goes for larger faults where the threshold may be > 0.3 . How did you determine an appropriate threshold, given that you also filter small faults

during the extraction phase?

Line 174: Should we not add the training time, model validation and QC, and labelling time as well to make the comparison fair? In this case the model was created specifically for this dataset, so the cost would not amortize over the application to many other datasets.

Line 180: The fatbox toolbox was mentioned earlier, do you need to repeat it?

Line 185: It is mentioned later that the fault score should not be equated with what I assume is a calibrated fault probability, yet here you mention that it is possible to determine how likely a fault is to occur. From a miscalibrated model this judgement can be misrepresented. See

Runhai Feng, Dario Grana, and Niels Balling, (2021), "Uncertainty quantification in fault detection using convolutional neural networks," *GEOPHYSICS* 86: M41-M48. <https://doi.org/10.1190/geo2020-0424.1>

Lukas Mosser and Ehsan Zabihi Naeini, (2022), "A comprehensive study of calibration and uncertainty quantification for Bayesian convolutional neural networks — An application to seismic data," *GEOPHYSICS* 87: IM157-IM176. <https://doi.org/10.1190/geo2021-0318.1>

for examples of how this can be ameliorated. Including validations against synthetic data and corresponding metrics (Mosser & Naeini 2022).

Line 188: Wouldn't you have to determine this empirically if your model does not differentiate based on strike, shape, or size? Some indications of this are the inability to detect faults that are oblique or parallel to inline crossline detection. In this case this is a given by design since the model is a 2D network. Validations using synthetic fault geometries would certainly help support or validate such assumptions.

Line 214: The statement "we can use (the fault score) as a proxy for how likely it is to encounter a fault" and the subsequent statement of "Bayesian neural networks ... able to predict true fault probabilities" are a bit contradictory. Would you agree that using the fault score as a proxy can only be done if the corresponding scores can be calibrated against independent data? I am not sure how you see the fault score being used in a quantitative manner, or whether you mean that the fault score can be used in a qualitative manner to indicate the presence of a fault which should not be mistaken for a true probability. Examples of such approaches are highlighted e.g. in Mosser & Naeini 2022 showing miscalibrations of U-Nets trained with balanced loss functions.

Line 221: Agreed, 3D fault extraction libraries would make for a great addition to the open-source software domain.

Regarding Figure 3, could the authors address why they choose to highlight only faults with a probability > 0.5 in the colour map, as opposed to figure 4?

Do the authors have any recommendations on how to choose filter sizes for the Gaussian blur? Is there a reason behind using a Gaussian blur to preprocess the fault score maps? Why would thresholding not be sufficient?

In figure 8, we can clearly identify some major faults not picked by the network and extraction in the lower right corner of the image. Is the seismic data that was used the same? It is addressed generally in the text, but I couldn't see if it was the same seismic dataset.

Have the authors considered processing the dataset in the main fault strike orientations i.e. NE-SW and NW-SE instead of inline direction? This could help better identification of oblique faults which are otherwise not well-imaged.

How was the Fault density square area measured, and how was the size of the averaging element chosen for Figure 10 D?