Response to Reviewer 1:

Comment on se-2021-23

Christopher Jackson (Referee)

Referee comment on "Fault Interpretation Uncertainties using Seismic Data, and the Effects on Fault Seal Analysis: A Case Study from the Horda Platform, with Implications for CO_2 storage" by Emma A. H. Michie et al., Solid Earth Discuss., https://doi.org/10.5194/se-2021-23-RC1, 2021

Dear Editor and Authors,

Thanks for giving me the opportunity to review "Fault Interpretation Uncertainties using Seismic Data, and the Effects on Fault Seal Analysis: A Case Study from the Horda Platform, with Implications for CO2 storage" by Michie et al. The general topic of the paper should be of interest to the readership of Solid Earth, dealing as it does with the subsurface structural analysis of normal faults and rift basins. It should also be of interest to people concerned with conceptual and interpretational uncertainty, both of which are critical considerations when working with structural data of any type. I have written numerous comments directly on the manuscript, which I have scanned and attached to this review. The numbers below refer to specific numbers in the manuscript.

1. "…how detailed the surface is…." could be replaced with "…how rugose the surface is…".
This has been altered, line 11.

2. It might also be worth mentioning; (i) the seismic line trend relative to fault strike; and (ii) the seismic or more accurately bin spacing in the original survey (which dictates the line or seismic line spacing available to the interpreter).
These points have been added, line 10.

3. Given that it is important and that it creeps into the paper Discussion, it might also be worth mentioning in the Abstract the detail vs. time trade-off.
This has been added on line 13.

4. Although a little pedantic, I would use "reflection" (i.e., the thing we observe and map in seismic reflection data) rather than "reflector" (i.e., the acoustic boundary or bedding surface that generates a reflection).
This has been altered, line 38.

5. What is the difference between a "linked" and a "composite" fault segment? They sound like the same thing to me. Maybe use one rather than both terms?
Yes, they are essentially the same thing ('linked or composite' was simply how it was phrased in the referenced article), so 'linked' has been removed, line 66.

6. There are some issues with the terminology you use here for the various fault growth models. I suggest you look at Childs et al. (2017) (https://sp.lyellcollection.org/content/439/1/1.abstract) and maybe use the terms therein, i.e., they refer to the "propagating" rather than "isolated" fault model. Granted, their Figs 1 and 2 are a little contradictory in terms of the terms they use, but I think the text (and subsequent papers by those and other authors) make this clearer.

Terminology throughout section 1.1 has been tightened using the suggested paper by Childs et al. (2017) – thanks.

7. Which other fault growth models are you referring to here? No references are provided to support this statement. I suggest you provide some or remove this statement.

Apologies, I must have missed this out. Reference added, along with more detail, line 72.

8. As noted by Childs et al. (2017) and Rotevatn et al. (2019), it is important to remember that the constant-length model does not preclude relay-breaching, i.e., it simply envisages that relay formation and breaching occurs relatively rapidly after fault initiation, such that a single, large-scale slip surface is formed prior to significant displacement accumulation. This was discussed in Jackson et al. (2017), more specifically in the last paragraph of section 1 (https://www.sciencedirect.com/science/article/pii/S0191814117300652?via%3Dihub). To truly distinguish between these fault growth models requires the integration of growth strata and not simply geometric analysis (i.e., the identification of displacement minima; see Jackson et al., 2017), something that is not undertaken or at least presented in your paper. As such, I am very hesitant as to whether your paper can really say anything about which fault growth model is most appropriate to the Vette Fault. Overall, although I think the paper can say lots of useful tings about the present fault geometry, in particular the potential location of now-breached, high-permeability relays, I am not yet 100% sure what you can say about how that geometry came to be in a temporal sense.
This is true that this current work doesn't perform a complete fault growth analysis, as this would be a separate study in its own right. Hence, I have modified this section, as suggested, to state that present fault geometry is analysed which shows the potential location of the breached, highly permeable, relays, lines 87-90.

9. As you discuss later, there is also the issue of failure to include so-called "continuous deformation" (folding) in the geometric analysis. This might be worth mention here, given it was again something we raised in Jackson et al. (2017) (the paper in response to the 2017 paper by Ze and Alves in JSG).
This has been added, line 91.

10. You could mention here that an outcome of relay breaching is the formation of abandoned hangingwall and footwall splays, which represent the now-deserted tips of the now-linked fault segments. These should be used in concert with simple fault bends to identify paleo-segmentation. In fact, do you see such things along the Vette Fault? It is noticeable that you do not show a single time-structure map in the paper, when such a map would be very useful in communicating aspects of the fault geometry (including, for example, the spatial distribution of folds, and the presence/absence of near-fault splays).
A Depth-structure map, along with a fault heave map of the top Sognefjord has been added (Figure 3).

11. You have already stated this a couple of times already. This could be removed without loss to the content of the paper.
Removed (line 102).

12. Precisely what about the hydrocarbon-water interface? The shape? The depth? Both? This is not clear, so I would encourage you to qualify this statement.
More detail has been added on line 129.

13. First, I am not sure I follow why larger faults have lower critical threshold values. Can you please outline why this is the case?

This was based on field and subsurface data analysis performed by Færseth (2006). However, this section on the shale smear factor has since been removed to shorten this section of the introduction.

Second, and more generally, there is a *lot* (c. 2 pages) of material in this general review of fault sealing, which delays the reading in getting to the study results. My question is whether all this material is needed to provide a basis for the analysis, results, and discussion that follow? If not, can some of it be removed without loss?
I've attempted to streamline this section (1.3), with several sections condensed or simply removed.

14. Could you include a regional cross-section to illustrate the geological setting of the study area better? At present it is really hard, from text alone, to get a good feeling for this. My comment about a time-structure map of, for example, the base syn-rift surface (which would record the cumulative rift-related deformation), also relates to this.
A regional cross-section has been added to figure 2.

15. What do you mean here by "good"? I suggest you estimate (using the extracted peak seismic frequency and the interval velocity at the depth of analysis) the actual seismic resolution. This will give the reader a better, more quantitative view of how good the resolution is, which I think is critical to convince people that variations in, for example, fault throw are really and not simply related to resolution-imposed picking limitations. Again, we discussed this in Jackson et al. (2017) when considering the very small throws (relative to seismic resolution) presented by Ze and Alves (2017).
Rough seismic resolution for the Sognefjord level has been added, line 199.

16. Do you mean "normal" in the context of the SEG standard? See here: https://agilescientific.com/blog/2012/4/5/polarity-cartoons.html.
No, European. This has been added (line 201)

17. It is a bit unclear what you mean here, i.e., did you not undertake such a QC for the other line spacings? Did you only do it when picking every 25 metres?
QC was done throughout – 'when every line spacing has been used' has been deleted.

18. How do you think a timeslice-based interpretation approach might have differed to your section-based approach? You say here timeslices were used to guide the interpretation, but if you do not see it as a valid option to replace the section-based approach, it might be worth stating precisely why not. For example, if variance timeslices are too noisy, compared to a standard cross-section, I think it would be valuable to say so. I say this because I think your paper has great value in guiding how people interpret seismic reflection data; in this case, also suggesting why *not* to use something is valuable too!
Why interpreting on timeslices has not been done has been added on lines 213-215. Seismic processing focused on resolving the Jurassic interval, as such the seismic quality is excellent at this location but can be significantly more noisy elsewhere than cross-sections. Hence, interpreting on timeslices alone would lead to huge ambiguity are not used for interpretation but for guidance only.

19. I am no expert in gridding approaches, so this section of text is a little hard for me to assess. However, it is not 100% clear to me from the text and Fig. 4 how the approaches vary in terms of what they're doing to the physically picked fault sticks, or

what their material impact on the fault geometry is. I guess this becomes clearer a lot later in the paper, especially through the use of Fig. 16.
It's not so much what they do to the fault stick, as much as what data from the fault sticks is honoured for each algorithm. I have attempted to make this clearer to the reader, lines 218-221.

20. A non-expert might not know what all these terms mean, especially fault stability and slip tendency, so I encourage you to include some supporting references…or provide a Supplementary Materials section containing information on how these various parameters are derived, what they mean, etc.
These have been described later (e.g. lines 257, 261, 264), so removed from this paragraph.

21. The term "polygon" (as used here and throughout the paper) is very confusing to me. I *think* you are using it in the sense that Badleys use it, whereas most interpreters will view it as a map-view feature that outlines the faut and essentially shows the fault heave at a specific structural level. Fig 5 does not currently really help me visualise what you mean by "polygon", so perhaps this and the text could be modified to make this clearer?
That's right – I suppose I never really thought of it as a 'Badleys' terminology, and simply assumed it was a well-known term. I have described polygons in a different manner (line 232), and changed the terminology to fault cutoffs throughout.

22. Again, this material related to the present stress state is not really my area-ofexpertise.
The in situ stress regime was provided to us from an internal Equinor report and from referenced papers (line 245 and 250).

23. Why do you say this, i.e., why might dip-linkage not be associated with a change in fault strike? Is this because the upper and lower segments are inferred to have the same strike prior to linkage? In any case, I wonder why you bring up dip-linkage when, later in the paper, this is not really something you talk about.
Since dip-linkage isn't something discussed, this has been removed.

24. What do you mean by "overlap" in this context? If the faults overlapped (prior to linking), then abandoned splays should be present. See my comments above.
Changed to breached relays, line 294.

25. Which structural level are you showing in Fig. 6? Top Sognefjord Formation? This is not clear in the text or figures.
Yes that's right – top Sognefjord. This has been made clearer in the text (290) and figure caption of Figure 6 (now Figure 7).

26. It is not clear to me why you need to normalise the plots in Fig. 7, given that the analysed length of the fault is the same in all picking strategies (i.e., 25 m to 800 m). Please could you clarify why this was done?
Length of the fault will vary with picking strategy as one end of the fault will remain the same, however the other end will decrease by the picking strategy size each time, due to the restricted size of the GN1101 survey. This explaination has been added to the caption for figure 7 (now Figure 8).

27. I am confused by your use of the term "should be" in this context. Why "should" these minima be picked in the 800 m sampling case? Also, I can't see the minima in the black

circles in the 800 m spacing plot. Or is the point that the black circles are only seen on the 100 m case and *not* the 800 m case?
That's exactly the point, that these minima aren't observed when using a coarse line spacing but are identified using a narrower line spacing. This section has been re-written to improve clarity, lines 311-314.

28. Could these corrugations also be associated with 'drift' in the picking by the human interpreter? Is this worth mentioning here? Oh, and I also would like to flag this interesting paper, which also reveals and discusses the origins of corrugations on seismic imaged normal faults:
https://onlinelibrary.wiley.com/doi/abs/10.1111/bre.12146.
Rigorous QCing was performed on the picking, to try to ensure continuity with the location of the fault pick, i.e. always on the footwall of the fault 'zone', so I would like to think that these corrugations are not caused by drift, so may in fact be due to the geology. Drift in picking is often what causes the irregularities/patches observed along the fault as it causes complications with triangulation of the fault surface.
Thanks for the paper suggestion – this has been referenced, line 319.

29. I am not sure I agree with this statement, given that the dominance of red and blues colours on the 'low-resolution' fault surface indicate that the fault still has an overall NS strike.
True, but there is a sudden change in strike, which is still observed when using the coarsest line spacing, whereas the subtle changes in strike are no longer observed. I've re-written this to improve clarity, lines 314-318 and 325-327.

30. Although you are not the first or only people to claim this, I have to say I disagree with the interpretation that corrugations are related to strike segmentation of normal faults. For this to be plausible, each segment would need to be relatively short (i.e., the corrugation wavelength, which in your case is a few hundred metres according to Fig. 8)…but very, very tall (i.e, the full fault height, which in your case is several kilometres, given the corrugations extend from the top to the bottom). This would result in faults with implausibly low aspect ratios (see Nicol et al., 1996) - https://onlinelibrary.wiley.com/doi/abs/10.1111/bre.12146. So, ultimately, I do not think picking strategy, "…may limit the interpretation of fault growth…".

I do share your skepticism about corrugations and fault segmentation. However, I don't believe these are simply a product of human error, but are likely to be geological, so I wanted to include strike in the fault growth analysis. Other suggestions to the cause of strike variation (e.g. from your 31. Comment) has been added, along with the fact that these corrugations would indicate potentially implausibly low aspect ratios, lines 344-349.

31. Or, perhaps, where fault bends occur as a result of out-of-plane propagation of the fault tips or complex early nucleation patterns? See https://www.sciencedirect.com/science/article/pii/S0191814106000320?casa_token=RfeVEhPMYFUAAAAA:97im5Akf7MWLjEkeslN3xtmWdRsR9RlNN-dpyDz1MG6GWnh7NejPQ0BzekKq3Fh5vZTyYFWEubg and https://www.sciencedirect.com/science/article/pii/S019181410600191X?casa_token=hBmvUeFBEXIAAAAA:e9K7bQaMHsUShAMzVWx7nv1TU8jsglkp5g9ekW95TKDzgYU3bYH4Lctm29GO711ROJyMhZrUkNA
This possible alternative cause for the corrugations have been added, lines 344-349.

32. Related to comment 25, which levels on the strike projection in Fig. 9 does the T-D plot come from? I ask because the black dashed lines are shown as vertical whereas the corrugations plunge slightly southwards, meaning they do not line up. Or at least they do not line up along the entire dip extent of the fault. Maybe they line up with the level at which the T-D plot is constructed, which is the reason to ensure that this is stated in the text, figures caption, and if possible labelled on the figure.
Thanks for this spot – fault cutoffs have been added to the fault plane diagrams at the top Sognefjord level, figure 10.

33. Regarding the comparative T-D plots for different picking strategies, maybe I'm misunderstanding something here, but shouldn't there be locations where the values are exactly the same? For example, every 32nd 25 m-picking line would lie up with an 800 m-picking line? If so, why is this seemingly not clear on Fig 10E?
You're right, however this plot as been normalized so that the length is kept constant (as the different picking strategy leads to slightly differing fault length). As such, the length of the fault picked on every line will be 775 m longer than that picked on every 32nd line, due to the restrictive size of the GN1101 survey.

34. Where is 0.4 on Fig. 10D? Do you mean values less than 40%? If so, there are hardly any values less than 40%. I'm a little lost here, so some clarification might be worthwhile.
Figure 10 (now Figure 11) is showing the calculated SGR, however this is based on a VShale curve and the throw – the value of 0.4 is the cutoff value for the VShale, not the SGR shown in the figure. Figure 10D is showing the calculated SGR values where low VShale overlaps occur at the Sognefjord level (which are all very high due to the high shale content in the overall succession). I've attempted to make the distinction between VShale and SGR clearer, lines 362-365.

35. Can you perhaps state how big these "patches" are and, most crucially, give a sense as to how they are distributed at and above the critical Sognefjord reservoir level? Surely, in the case you present here, the variability in these locations, arising from the various picking strategies, is what's really key?
These details have been added, lines 389-390. However, it is unclear and unpredictable where the fault will fail given the increased pressure from injected CO2 (if the fault is in critical failure), so these 'patches' may not neccessarily indicate where failure may occur (lines 463-464).

36. Why do you propose 100 m? What do you not believe it real, geologically speaking, at spacings less than 100 m? And/or what do you not think is important, in terms of CO2 storage and potential leakage, at these very small scales?
This has been removed from this section, as I have proposed and discussed the optimum picking strategy as 100 m in the discussion section (lines 451-456). This proposed picking strategy is based on the fine balance between human error and surface generation (triangulation) – shown best in figure 15. Any subtle variation between two adjacent lines (which is almost unavoidable due to the scale of seismic resolution, and is a product of human nature, despite thorough QC) will create a highly irregular fault surface. This is particularly true when using a triangulation method that honours every point. It may seem counter intuitive to suggest not to use every line, particularly when faults are very heterogeneous and irregular in nature, however this is a product of the scale of analysis used in seismic studies. This study attempted to highlight the need to pick according to an 'optimum' strategy, whereby inherent irregularities can still be captured, which are not (or

less so) a product of human error or triangulation, but not overly smoothed out when using a very coarse picking strategy.

37. This whole section is a little 'wordy' and I think the language could perhaps be streamlined and simplified a little There are many, many occurrences of the terms "dilation and slip tendency"; could these perhaps be used only when essential?
I have re-written this section (4.3.2 fault stability) so it is more streamlined.

38. Rather than stating "lines" in terms of seismic line spacing, could you stick with the terminology used throughout the paper up to this point, i.e., horizontal spacing of 25 m, 100 m, 800 m, etc.
Terminology changed to metres to keep consistency throughout.

39. What do you mean by the inlines and crosslines "not tying precisely"?
Ah, I meant the *interpretations* made on crosslines may not tie precisely with the intersecting inlines. This has been re-written to improve clarity, line 465-468.

40. Is it not rather obvious that a relatively thin reservoir, in a relatively muddy sequence, next to a relatively large fault would lead high SGRs and a high likelihood of fault sealing? In that sense, I do not find it surprising that the picking strategy was an important control on this.
Yes, this is true. 'Surprisingly' has been removed (I don't really know why I put it in!).

In summary, this is a very interesting and important work, which I very much enjoyed reading. Overall, the paper is well-structured, well-written, and the work is thorough and, for the most part, convincing. However, as I hope is clear from my comments above, some revisions will, I hope, help further improve the manuscript. In general, the English and grammar are very good, although there are a few places where these could be improved. I am more-than-happy for the authors to contact me to discuss any of the issues raised in my review.
Yours sincerely

Response to Reviewer 2:

Dear authors,

I enjoyed reading this manuscript which highlights an under-investigated and under-reported part of subsurface geological modelling, and deals with subjective interpretative strategies and decisions. The paper is very well written, structured, the figures are clear and easy to understand. The paper does a very good job investigating the uncertainty presented by different picking strategies, however I think the discussion can be improved by considering how this particular uncertainty compares to other uncertainties present in seismic inversion, interpretation, and geological model making. If we consider all sources of uncertainty in a classical error propagation framework, the source of the largest error is the most important, while identifying error variability in lower error sources becomes insignificant. I think this paper presents cases where the uncertainty presented by picking strategies is the dominant source of uncertainty and therefore important, as well as those

where it may not be the leading source of error. My further comments below are mostly around this theme and specific to line numbers.
Unfortunately, critically assessing all the uncertainties in the entire seismic interpretation workflow to identify the source of largest error is beyond the scope of this manuscript (but would be extremely useful!). For example, I think it would be very hard to say whether human error / interpretation bias has a more important control on the framework model produced, or line spacing. Moreover, we do not have enough data to assess uncertainties pertaining to human error or seismic resolution. However, we have discussed other uncertainties and errors throughout (e.g. seismic quality, human error, interpretation bias etc), and all these sections have been expanded on within the discussion to highlight that there are other uncertainties than just line spacing. E.g. lines 503-508, 523-529.

Lines 195-200 Could the authors please comment on the resolution of the velocity model used in generating the seismic images? Was it generated using tomography or full waveform inversion? Was any lateral smoothing applied to the velocity model before processing and imaging? Because, although the seismic itself is binned on 25 m x 12.5 m grid, using a velocity model which was not also calculated independently in each bin (e.g., by tomography), or using a model which has some form of smoothing applied to it, introduces an error or lowers the effective resolution of the final seismic image such that interpreting at 25m becomes overfitting or at least possibly unnecessary. This is important for further discussion, where in some instances (for example reactivation potential plots) the closest possible picking (25 m) might produce the best fit for the seismic, but that might not be equal to the best fit for the underlying geology.
The velocity model was created using TD curves in order to do a simple depth conversion. This is likely to cause some problems with the images due to the high velocity contrast at the basement level, however at the level of interest (the Sognefjord), the velocity is well constrained, as such the depth converted seismic should be suitable for this analysis. No processing has been performed. A comment concerning the resolution of the seismic simply using the frequency and interval velocity at the depth of analysis (Sognefjord level) has been added, lines 198-199.

Lines 355-360
I think the finding of this section that SGR does not vary significantly irrespective of the picking strategy is a positive finding, and indeed will be positive for those using lower resolution seismic or legacy data. I would be cautions suggesting that coarser picking slightly overestimates the SGR data (Lines 359-360). SGR in itself is of course an approximation with an associated uncertainty (the content of clay in particular, from gamma ray log, is a source of larger uncertainty), and I think the uncertainty of some of those assumptions is higher than that presented by the picking strategy.
I absolutely agree with this. The SGR section within the results has been re-written, and a section within the discussion has been added to highlight other uncertainties within this analysis, lines 557-559.

Line 396 I think the framing of this as 'correct' and 'incorrect' assumptions is slightly misleading. I think we are trying to find an approach that provides the best fit to data, but the best fit might not be 'true' in any case as in cases were we are within the limits of the seismic resolution, we may be still within an uncertainty of the velocity model used.

It is true that we will not know what is ‘correct’ or ‘incorrect’ within the subsurface. I have changed ‘incorrect’ to ‘unlikely’, line 402.

Lines 452-453 I think here the paper is arriving close to the lowest possible error that can be used as base error. The relative differences in geometry produced by picking by two skilled and local geology-informed operators is the type of error that in practice currently probably cannot be further minimised, especially across the community. I would view this amount of variation as base error. Looking at figure 15, if we were to smooth the fault surface until a point where dilation tendency plots on A and B look the same, I think this is the lowest possible error we can get down to. The question is then the following - what is the lowest picking interval that can be used that doesn't further decrease this resolution? I think the suggested 100m is a likely outcome, but I think this concept needs a discussion.
Smoothing by increasing line spacing is likely to still retain some discrepancies between the interpreters (as all lines and points of a picked segment will never be exactly the same). Smoothing using the gridding method with additional smoothing may create similar fault surfaces, however this won’t provide information concerning the lowest picking interval that doesn’t decrease the resolution. I have added a small section discussing suggested picking strategy may vary with human error, line 460.

Lines 460 - 465. This paragraph is an important discussion of over-fitting and I recommend expanding it. I think in many cases a degree of smoothing brings the model closer to best fit, depending on what is being modelled. For example, highly irregular fault surfaces will display a great range of fault reactivation potential values (Fig 12) over a surface area. In practice, we are likely not expecting a fault to be reactivated just in one of these spots or in a number of identified spots without affecting areas in-between. In this type of plot, a higher degree of smoothing seems beneficial to me (100m and above) vs 25m which produces a lot of scatter.
Thanks for the suggestion – this section has been expanded on (lines 461-465).

Lines 509-511 I'm not an expert in differences between triangulation methods, but here it would also appear that if all three methods are equally commonly used, then the difference between them is the base error. If there are advantages in using one or the other in certain cases, then that can be discussed.
I don’t think there are ‘advantages’ to using one over the other, but simply that each one contains different levels of smoothing, or ‘best fit’ to all data points. I wouldn’t want to recommend which triangulation method is better than others, instead I have highlighted what each method does and what the results are, so the user has a better ability to choose the method. This section has been expanded on, lines 517-520.

Lines 557-564 Here I think the authors do a good job discussing how other uncertainty sources (cohesion) are likely to be more significant.

Overall, I think this is a great contribution and a very well organised and presented case-study example. I would recommend for publication after these minor revisions.