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.