Response to Reviewer 2 – Anonymous

We thank the reviewer for their comments (repeated below in black) and provide detailed responses below (in red).

First of all, I would say that the authors are top scientists in this field and accordingly, the idea and the methodology reported in this paper seem to be very promising. Moreover, for people like me with a prevalent geological background, the pure statistical part of the paper can be hard to be read just because of the background.

Thanks. One key aim of our ms (see lines 52 – 59) is to explain the underlying theory and statistical background to the Response Surface Methodology for just these reasons. And according to Reviewer 1, it is "well written, carefully explained and thoughtful".

However, the geological data seems to be, in my opinion, poorly exposed here and the statistics are sometimes completely detached from the geological data making this paper quite difficult to be read from a Solid Earth reader.

We don't understand what is meant by "geological data seems to be ... poorly exposed here". We have used the publicly available geological data for each case study, and cited all the sources. Also, we do not understand the comment "the statistics are sometimes completely detached from the geological data". In the absence of complete certainty in the available data, we have used specific statistical distributions to model the consequences of uncertainty.

Generally speaking, the paper faces a very interesting problem, and the method is innovative and very exciting. As far as I can see the methodology is new and for this it must be tested and verified yet. The authors attempt to do this by presenting two case studies with the aim to show "how combined RSM/MC approach can be used to estimate the probability of slip on one or more faults".

We agree that this is interesting, innovative and exciting.

However, the two cases are not very well constrained in terms is of boundary conditions making the probability estimation quite confused.

We do not understand what the reviewer means by "boundary conditions". We are not conducting a numerical modelling analysis of a fixed spatial or temporal domain, e.g., of tensor fields or conservation equations using finite differences, and therefore the notion of formal boundary conditions is misplaced, in our opinion.

Our analysis, described in the first two sections, focuses on modelling the consequences of uncertainties in all of the possible input parameters involved in the quantification of fault stability (using either fracture susceptibility and slip tendency). As such it is a direct extension and development of the work presented by, for example, Chiaramonte et al. (2008) and Walsh & Zoback (2016). We do not think the probability estimation is "quite confused" (cf., comments by Reviewer 1).

Moreover, the two performed analyses (Porthtowan Fault Zone in Cornwall, UK and Coalfields in South Wales and Greater Manchester, UK) differ in so many aspects and, more importantly the presented results are different in terms of delivered outputs. This make the reading quite confusing and at the end of the paper I got lost about the point that the authors would like to address. In my opinion to test a new methodology we should apply this in areas where data are known as much as possible to see if the model prediction are reliable. In this case since the two areas are poorly constrained, this exercise is difficult to be followed and the results even more difficult to be understood.

We agree the case study areas are different, and the chosen modelled outputs are also different. This is all deliberate. Our intention is to demonstrate the scope of the method (combined RSM and MC) to make useful predictions about fault stability in terms of fracture susceptibility (United Downs) and slip tendency (coalfields) in the face of uncertainty.
As noted above in Response to Reviewer 1, we will remove the Manchester coalfield case study to reduce the length of the ms. We hope this makes it easier to appreciate the differences – and more importantly, the value in those differences – in the two case studies.
In relation to "we should apply this in areas where data are known as much as possible to see if the model prediction are reliable": we know of no such datasets. In the case of United Downs – arguably one of the best constrained sites involved in geothermal energy – all of the data remain uncertain (to varying degrees), and this is one of our key points: even for areas with apparently "good" data, we argue that the existing uncertainties are significant and have consequences.

The discussion paragraphs more than discuss the results present a list of what we should know to better assess the seismic risk and the main message seems to be that we would need to know a lot of things. I can kind of agree with this but, once again, this makes the main message of the paper more confused.
We disagree with this comment and agree with Reviewer 1 that the ms is "well written, carefully explained and thoughtful".

I strongly suggest the authors to simplify the paper in two ways.

1. Try to organize a sort of sensitivity analysis of the involved parameters in a more structured and ordered way in order to facilitate the reader
2. Focus in one area and compare the results with something actually observed.

For the first point, sensitivity analyses are already included in the worked examples and in the case studies; for example, we use CDF plots to explore the absolute sensitivity to selected parameters and we use tornado plots to rank the relative sensitivities (see Figures 4, 5, 7 & 8).
For the second point, we think the reviewer might have missed the point. We know of no site or area where the observations are known perfectly, i.e. with 100% certainty.

I think that we all agree that there are many topics related to the risk assessment (fault length, roughness, friction, fluids, background seismicity, regional strain rate, and many many others) but in doing this exercise authors must clearly state the assumption and critically analyze the results. In this paper I had the impression that speaking about the many variables we lose the point of the paper, I would say that sometimes less is more.
We have stated the assumptions used throughout (e.g., Mohr-Coulomb failure), and we critically analyse the results through detailed statistical analysis of the outputs. One of our main aims, clearly listed in the Introduction, is to provide a clear and detailed explanation of the method (in our opinion, so far lacking in previous publications using similar methods). This entails some detailed and "careful explanation" (Reviewer 1).

Minor points:
I am not so convinced about the statistical discussion that is sometimes too focused on the pure statistics and few on the geology behind. For example, can we find a geological meaning to the "asymmetrical or skewed" distribution of some parameters?
This is one of the issues raised by our ms, and clearly discussed! By trying to accommodate the fact of uncertainty in all input parameters – stresses, orientations and rock properties – we are faced with making choices about the nature (shape) of their distributions. We clearly state that there is

currently insufficient published data for many of these parameters – especially some critical ones such as cohesion and friction – to find any “geological meaning”.

I Am not expert on Response Surface Methodology (RSM). However, the paragraph Statistical analysis of geomechanical fault stability start with a discussion on the governing equations for RSM following a quite long description that ends with the definition of Ts by meaning of the very well-known direction cosines (e.g. Ramsay and Lisle 2000). In other word I can’t really see why the authors need introducing the RSM theory to infer the Ts definition.

The reviewer has perhaps missed the point. We are not “inferring” the Ts definition. The equations for Ts are given in their full format (i.e., in terms of direction cosines) to highlight one of the key issues: there are 8 input parameters, and they are all, in general, uncertain. This is picked up in the succeeding paragraph (line 221 in the original ms). We need to show the full equation for Ts before we make this crucial point.

A lot of acronymous BGS, CDF, are used but not defined. Even if they are quite easily understandable, this gives the impression of a lazy writing

We presume the reviewer means “acronyms”. BGS is the British Geological Survey – we will add a definition for that. CDF is defined on first use, on line 138.

The discussion on the relationship between fault length and events magnitude starts with this and ends with discussing the relationship between fault length and number of events. I would say that the two (maximum magnitude and number of events) are surely correlated but they are not the same thing.

Agreed. But we do not say they are the same thing.

Line to line comments:

Line 228 I would say that fluid pressure also influences Ts (e.g. De Paola et al., 2007)

We strongly disagree. Pore fluid pressure plays no part in the formal definition of slip tendency (Ts) – see Morris et al., 1996. Moreover, the influence of pore fluid pressure on the potential for failure is better understood in terms of fracture susceptibility – i.e., the pore fluid pressure increase needed to drive the stress state on the fault to failure (Streit & Hillis, 2000).

Line 239 is CDF the cumulative distribution function? Authors should state this somewhere.

Yes it is. It is defined on first usage, on line 138 of the original ms.

Line 326 alfa has been not defined

Definition for alpha (α) will be added.

Line 698 Why these may be the ones most likely to slip?

We are highlighting the *possibility* that unmapped (i.e., unknown) faults *may* be most likely – due to all the factors discussed in this paper. The point is about unmapped faults, or so-called “known unknowns”.

Line 700 Some of this “mismatch” could be explained by the dip of the faults measured at the surface, but not all. What the author mean here?

We mean that the surface traces of the faults shown on our maps may not coincide with their extension at depth, e.g., for faults that dip at less than 90 degrees. This could explain some of the apparent mismatch between the recorded earthquake locations plotted on the map relative to the surface traces.

Line 742 The observational record shows that bigger fault zones. I would say that there are a lot of physical reasons behind this. Moreover, empirical relationships such as those suggested by Wells and Coppersmith 1994, or Leonard 2010 should be cited here.
Thanks for these suggestions. We will add these papers.