Abe & Deckert produce fracture surfaces in DEM models and laboratory experiments on sandstone and limestone. They measure the roughness of these surfaces with a variety of metrics. They find an insignificant influence of confining stress on fracture roughness, and self-affine numerical and laboratory fractures. One key finding is that biaxial loading produces greater roughness than triaxial loading. However, the manuscript does not adequately explain why this difference in loading would produce the difference in roughness. Throughout the manuscript, there are a few other interesting results that would benefit from additional mechanical explanation. For example, comment #9: why do the natural rock samples have different ranges of roughness in the slip-parallel and -perpendicular directions?

I suggest that this manuscript may be published after the authors provide more thorough explanations of their results. More detailed comments follow below.

Best,
Jess McBeck

1) Line 115: "three of the 47 equations presented there have been chosen". Why did you select these equations?
2) Line 202: "These parameters do not provide a direct match to the mechanical properties of the rocks used in the laboratory tests (Section 3.2), but the important ratio between failure strength of the material and the confining stress applied in the laboratory experiments lies well within the range covered by the numerical models (Fig. 5b)." Was there a particular reason that calibration of the models did not produce the precise values of the Young's modulus and UCS of the lab? Or was the problem that the lab USC ranged from 285 MPa to 80 MPa? Also, the Young's modulus of the lab rocks does not seem to be mentioned in section 3.2. And why is the ratio between the failure strength and confining stress more important than the absolute values of the UCS and Young's modulus? I would imagine that the stiffness of the rock has an impact on the way the rock fractures, and the resulting fracture roughness.
3) Line 224: "3D point cloud data with c. 2.2 mio data points". Presumably mio indicates million here? It could be good to change this abbreviation.
4) Line 222: "Only in one experiment with a confining pressure of 30 MPa post-deformation fragments were large enough for our planned fracture surface analyses (Fig. 4b). " If I understand correctly, the lab experiments yield only two fracture surfaces for which you can calculate the roughness, one from a sandstone and one from a carbonate. If this is the case, then please state it explicitly. Otherwise, it could seem that you have several natural fracture surfaces (i.e., Figure 5).
5) Paragraph at line 237: Why were these estimates outside the range of acceptable values? Could it be related to the inherent roughness of the particles? I wonder if you ran simulations with a larger (0.1-1.0) or smaller (0.3-1.0) particle radii range you would see differences in the calculated roughness, with lower roughness for the larger particle radii range? Later, you show the difference between the height-height correlation functions of a numerical fracture and an arbitrary cut through the model, but did you do a similar calculation for the roughness metrics described in this earlier paragraph? It could be useful to see how the roughness metrics change when you change the particle size distribution, although this would require changing some of the other model parameters, such as the timestep size etc.
6) Line 259: "is possibly at least in part an artifact of the different size of the fracture surfaces between the two model groups" Why is it the case that the size of the fracture surface correlates with the RMS such that larger fracture surfaces would have larger RMS? I understand your explanation of how the size of the fractures differs between the unconfined extension and compression cases, but I do not see why larger fracture surfaces necessarily lead to larger RMS. One may expect greater variations from the mean from a smaller surface as there are less data points producing the mean.
7) Line 319: "The results show that the estimated JRC is dependent on the sampling resolution, i.e. the number of sampling points on the profile, specifically that the calculated value of the JRC is increasing with smaller sampling intervals". Why did you not observe this influence for the numerical models when you calculated the roughness parameters at parallel and perpendicular profiles? And if this is a known effect, then why not shorten the profiles along the dip direction so that they match the length along the strike direction? It would be

interesting to know/constrain if the observed anisotropy is real or not. For example, see the next comment.

8) Line 246 "The results did show that the mean estimated JRCs for the profiles differs by less than 10% between the two direction, which is generally less than the standard deviation between the profiles within one direction" and Figure 15: The laboratory data shows some anisotropy in the roughness, whereas the numerical models do not. Why is this the case? In the discussion, you mention how the indestructible particles and lack of breakable grains could contribute to this point. I also wonder if the faults in the experiments experienced slightly more shear displacement than the fractures in the numerical models. In the numerical models, it's straightforward to select the timestep when the fracture first breaks, and before it slides. But in the lab experiments, I imagine that the deformation could not be stopped at the exact moment of fracture, and that some (perhaps small amount of) slip must occur after the fracture forms (at least for the experiments with compressive loading conditions). This slip could then help produce the anisotropy of the natural lab fractures. In addition, could you also measure the roughness of the numerical fractures after they slip, i.e. some timesteps after the fracture forms? It would be nice to see how the roughness evolves with slip. I wonder if the influence of confining stress on roughness would be larger (not insignificant) when the fracture surfaces slip.

9) Line 333: "Therefore the best estimates for the average JRC of the fracture surfaces produced in the laboratory experiments are for the sandstone $J RC \approx 9 - 11$ in the direction parallel to shortening direction in the deformation experiment and $J RC \approx 11 - 13$ perpendicular to it (Fig. 15). For the limestone the estimates are $J RC \approx 6.5 - 7.5$ in the parallel direction and $J RC \approx 16 - 17$ in the perpendicular direction. " Why is the limestone rougher than the sandstone in the slip-perpendicular direction, and less rough in the slip-parallel direction? It's hard to see how this result is directly tied to grain size. Maybe the calcite grains in the limestone are more mechanically anisotropic in strength than the sandstone quartz grains?

10) Line 375: "Based on the results from the numerical models there appears to be a trend towards higher roughness for fracture surfaces generated under transversely isotropic stress conditions, i.e. standard triaxial compression ($\sigma_1 > \sigma_2 = \sigma_3$) compared to those generated under true triaxial conditions ($\sigma_2 \neq \sigma_3$). " Why is this the case? Maybe the orientation of the fracture with respect to the intermediate and minimum compressive stresses could provide insights? How was the fracture oriented relative to the horizontal axes (sigma_2 and/or sigma_3)? Why do smoother surfaces form under true triaxial conditions?

11) Line 419: "Looking at the relative timing of bonds breaking suggests that the fracture surfaces in the DEM models grow by coalescence of micro-cracks despite having a Hurst exponent closer to 0.4." Although the focus of this paper is the fracture surface, it would be useful to see how the fractures coalesce to form this plane, i.e., the position of the fractures (broken bonds) in the timesteps leading the final failure. I would also be interested to see the timeseries of fracture positions in models with true triaxial loading and biaxial loading conditions.