Reply to Reviewers Comments

Text in *italics* are reviewers comment, normal text is our reply. The reference section only contains those references which are newly introduced into this reply, i.e. which are not already in the manuscript.

Reviewer 2

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?

We did choose these 3 equations because they all show a high correlation coefficient, 0.986 for JRC_1 (Eq. 5), 0.951 for JRC_{31} (Eq. 6) and 0.971 for JRC_{34} (Eq. 7) [Li & Zhang, Table 2] and because we wanted to use 3 equations based on different underlying roughness measures. JRC_1 is based on the Root mean square of the first deviation of the profile as defined in [Li and Zhang, 2015, Table 1], JRC_{31} on the Roughness profile index and JRC_{34} on the Profile elongation index.

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?

It would have technically been possible to calibrate the DEM material to a UCS of 285MPa. The reason that this wasn't done is that, due to factors outside our control, the laboratory experiments were performed after the bulk of the numerical modeling had been done. So the DEM material was calibrated to typical values for the rock types used as described in the literature. Unfortunately it turned out that the limestone has an unusually high strength.

Also, the Young's modulus of the lab rocks does not seem to be mentioned in section 3.2.

The missing values (48GPa for the limestone, 12.5GPa for the sandstone) will be added.

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.

We would argue that indeed the ratio between strength and stress is the more important factor. The reason is that fracture criteria can typically be expressed in a non-dimensional way if they are formulated in terms of stress-strength ratios and that many aspects of fracturing are determined by where on a failure envelope the stress state is located under which a particular fracture is formed.
However, we would also agree that rock stiffness will have an influence on the way the rock fractures. This applies in particular because the ratio between stiffness and strength will determine the failure strain. The unconfined failure strains of the calibrated numerical models and the rock samples differ by only a factor of 2, i.e. the DEM material fails at a strain of ~$3*10^{-3}$ under uniaxial compression, both rock types at about $6*10^{-3}$. We did run a few models with higher strength during the calibration phase which did not show any obvious differences.

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.

Will be changed.

*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,*We did *then please state it explicitly. Otherwise, it could seem that you have several natural fracture surfaces (i.e., Figure 5).*

We will modify Figure 5b to make it clear that only two of the laboratory tests were used. See panel (b) in figure below.

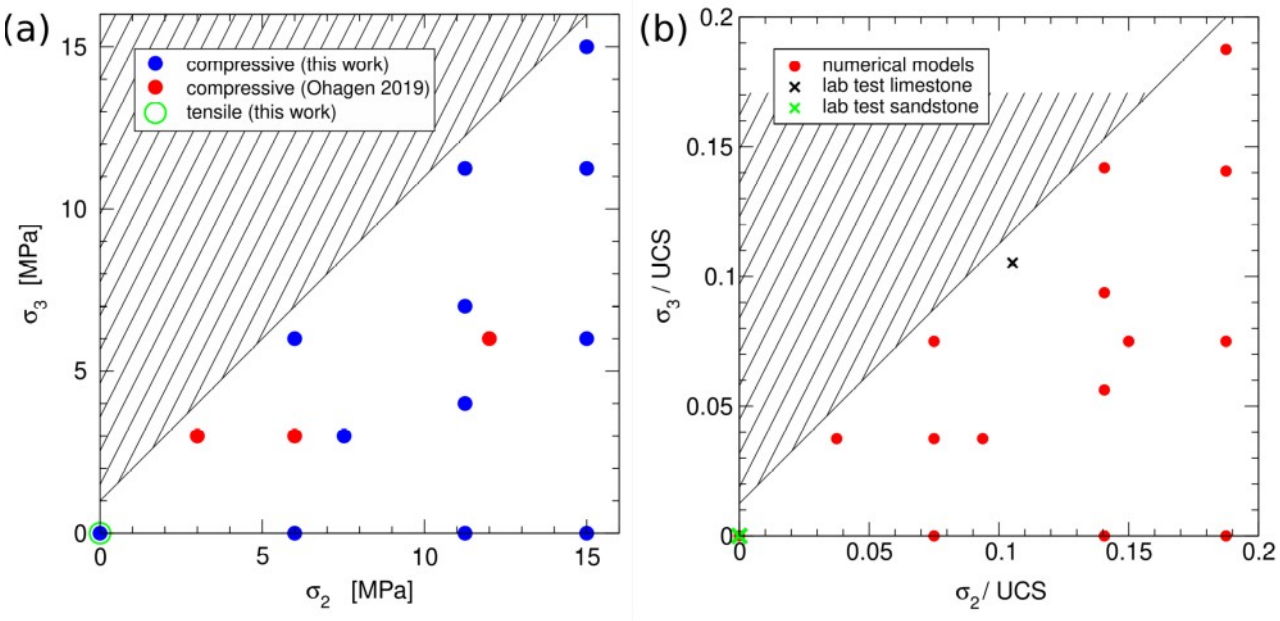

Figure 5 (modified)

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?

There is a contribution of the intrinsic particle roughness to the JRC, but as discussed in lines 346-356 it is too small to explain the very high JRC values calculated for the numerical models.

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.

We did not run models using a narrower range particle sizes, such as 0.3-1.0. However, we did run 2 small sets of simulations with a wider range of particle sizes (0.15-1.0 and 0.1-1.0). The sets consisted of 5 simulations each, all performed under true triaxial conditions using σ_2=6MPa and σ_3=0. Results did not show a statistically significant difference in Hurst exponent or JRC compared to the equivalent simulations performed using a particle radius range of 0.2-1.0. See table below.

particle size range	Hurst exponent	JRC	JRC anisotropy
0.2-1.0	0.414±0.5	26.1±1.8	2.0%
0.15-1.0	0.415±0.85	25.4±2.4	3.2%
0.1-1.0	0.398±0.96	24.2±2.3	0.4%

We will add this data and the explanation above to the discussion section of the manuscript.

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.

Given that the fracture surfaces are self-affine, it would be expected that the variation in "height" between points on the surface scales with their distance as shown in Fig. 9 in the manuscript. Therefore a larger surface would be expected to contain larger height differences, resulting in a larger RMS roughness.

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?

The JRC data for the numerical models are calculated using a constant sampling interval and therefore a variable number of sampling points depending on the profile length. (see also next comment below).

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.

The roughness measures underlying the JRC calculations ("Root mean square of the first deviation of the profile", Eq. 9 in the manuscript) are local measures which depend only on the differences between neighboring sample points. The calculated JRC values do scale with the size of the sampling interval, i.e. the distance between sampling points, not the total number of samples. So “shortening” of the dip-direction profiles in the sense of clipping them to the same length and number of sample points as the strike-direction profiles might change the JRC of individual profiles, but it does not affect the overall statistics.
Additionally, at least for the limestone the observed anisotropy is much too large to be explained by sampling effects. The ratio between the JRC calculated for dip- and strike-direction profile using the same sample point distance differs by a factor of at least 2 for all sample point distances tested (Fig. 15 in the manuscript), which is significantly larger than the observed change of the JRC for difference sample intervals (same Figure).

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.

As stated in the reply to reviewer 1 (page 1, lines 46-51), the grooves in the limestone sample, which are largely responsible for the high anisotropy measured there, do not appear to be abrasion features.
To look at the roughness evolution of the fracture surfaces with increasing deformation of the sample we did perform a small number of simulations which did not stop immediately after the formation of the fractures, but instead continued deformation to a total axial strain of up to 12%. This is significantly larger than the strain occurring in the laboratory experiments, where total axial shortening did not exceed about 2%. In particular, the amount of shortening occurring after the peak axial stress was reached, i.e. after failure, was generally less than 1%.

For the surfaces extracted from those simulations, we did not observe a significant change of the Hurst-exponent with increasing shear offset of the surfaces (see reply to reviewer 1, pages 1-3). For one of the models we did also calculate the JRC of the surfaces at various stages of the simulation (see Figure below). The data shows that there is also no significant change of the JRC for the shear offset considered in this model, which would be equivalent to ~1cm in the laboratory samples, and under the conditions of this model (true triaxial stress, σ_2=7.5MPa, σ_3=3MPa).

This seems to confirm again that under the small shear offsets relevant for our experiments, there is very little evolution of the surface roughness, at least as far as it concerns the roughness parameters calculated here (Hurst-exponent, JRC). In particular, the data would suggest that any effects due to small, but the non-zero, shear offset in the laboratory experiments are much too small to explain the observed differences between numerical simulations and laboratory experiments.

We will add a paragraph discussing this issue to the “Discussion” section of the manuscript and the figure below, and also Figures 1 and 2 from the reply to reviewer 1 (pages 2,3) to the supplementary material.

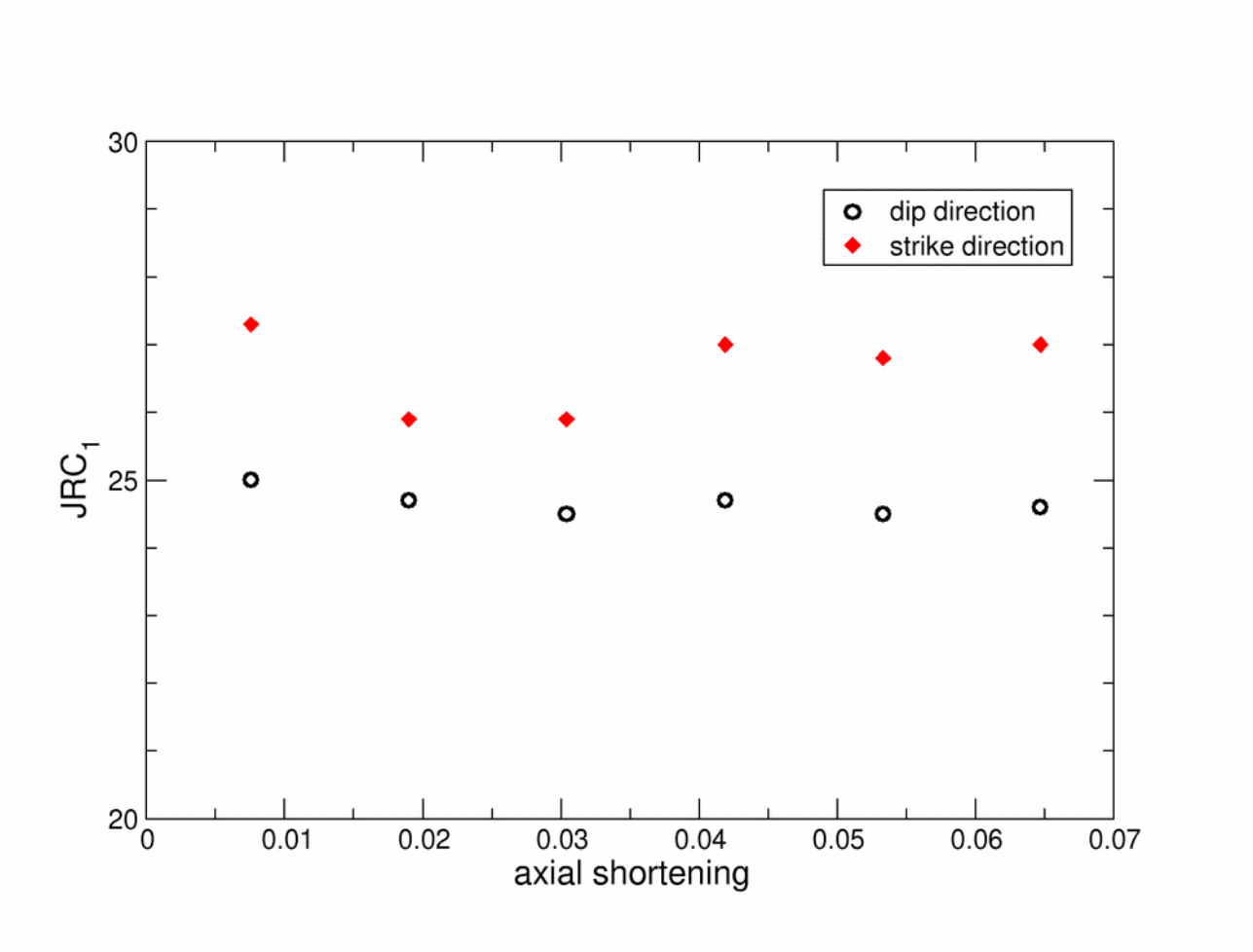

Figure 1: Evolution of the JRC with increasing shear offset of the surfaces. Data points are averages of the two surfaces of the same shear fracture.

9) Line 333: “Therefore the best estimates for the average JRC of the fracture surfaces produced in the laboratory experiments are for the sandstone JRC ≈ 9 − 11 in the direction parallel to shortening direction in the deformation experiment and JRC ≈ 11 − 13 perpendicular to it (Fig. 15). For the limestone the estimates are JRC ≈ 6.5 − 7.5 in the

parallel direction and JRC ≈ 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?

The high roughness in the perpendicular direction in the limestone sample is mainly due to the large longitudinal grooves on the surface visible in Fig 4b and Fig 13a. The mechanism responsible for the generation of those structures is, unfortunately, unclear to us.

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?

This is an interesting idea which could probably be tested in future work. However, we do currently have no indication that the calcite grains in the limestone show any sort of preferential alignment, which, one would presume, should be necessary for the grain scale strength anisotropy to have any effect much beyond the grain scale. Unfortunately all limestone samples were taken in the same orientation with the long axis of the cylindrical samples, and therefore σ_1 during the triaxial tests, perpendicular to the bedding of the limestone, so we do not know if the macroscopic strength of the limestone is anisotropic or not. However, SEM images of limestones sampled from the same location appear not to show any preferred grain orientation.

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 (σ1 > σ2 = σ3) compared to those generated under true triaxial conditions (σ2 ≠ σ3). ” 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)?

The fracture orientations were as one would expect under the stress conditions. The dip angle was within 25-35° of σ_1, i.e. 55-65° assuming σ_1 to be vertical. The strike was usually within ~10° of σ_2 in the true triaxial models (σ_2 != σ_3) and more or less randomly distributed in the transverse isotropic conditions ($\sigma_2 = \sigma_3$).

We will add this sentence to the “Results” section of the manuscript.

Why do smoother surfaces form under true triaxial conditions?

Unfortunately we can only speculate about the reasons at this point. One, purely speculative, idea might be that, if we assume that the through-going fractures which we analyze form by coalescence from smaller, precursory, fractures, those precursory fractures have their strike angles constrained to a narrow range if σ_2 != σ_3, but that there is no such constraint if $\sigma_2 = \sigma_3$. If this is the case, than the coalescence of those fractures might lead to smoother large-scale surfaces if they all have similar orientations. Unfortunately the numerical models used in this work do not have the resolution necessary to study this process.

We will add a paragraph about this to the “Discussion” section of the manuscript.

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.

While the details of the evolving micro-crack distribution prior to the main failure of the sample have indeed been beyond the scope of the work presented in the manuscript, we did store sufficient snapshots from a number of models to reconstruct time series of micro-crack distributions where every "micro-crack" visualized is a single particle-particle bond breaking. See below a set of snapshots from a biaxial ($\sigma_2 = \sigma_3 = 6$MPa) and a triaxial ($\sigma_2 = 15$MPa, $\sigma_3 = 6$MPa) model. Timing of the snapshots is shown on a plot of axial stress over time below the snapshots. Individual micro-cracks are colored by failure time.

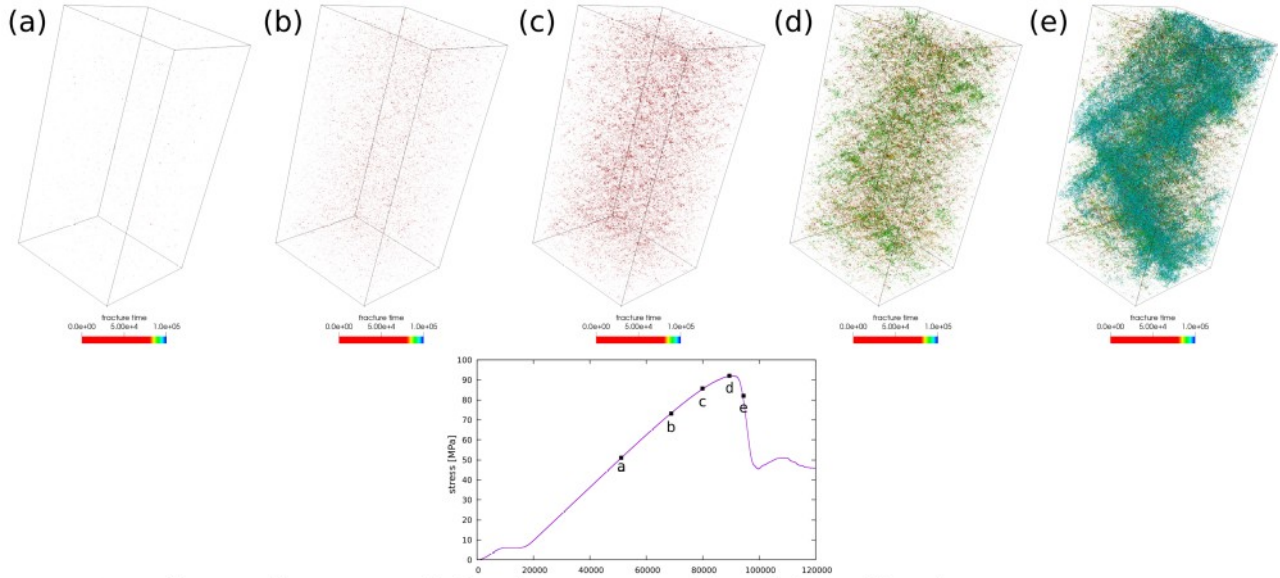

Figure 2: Evolution of micro-crack distribution in model with biaxial loading

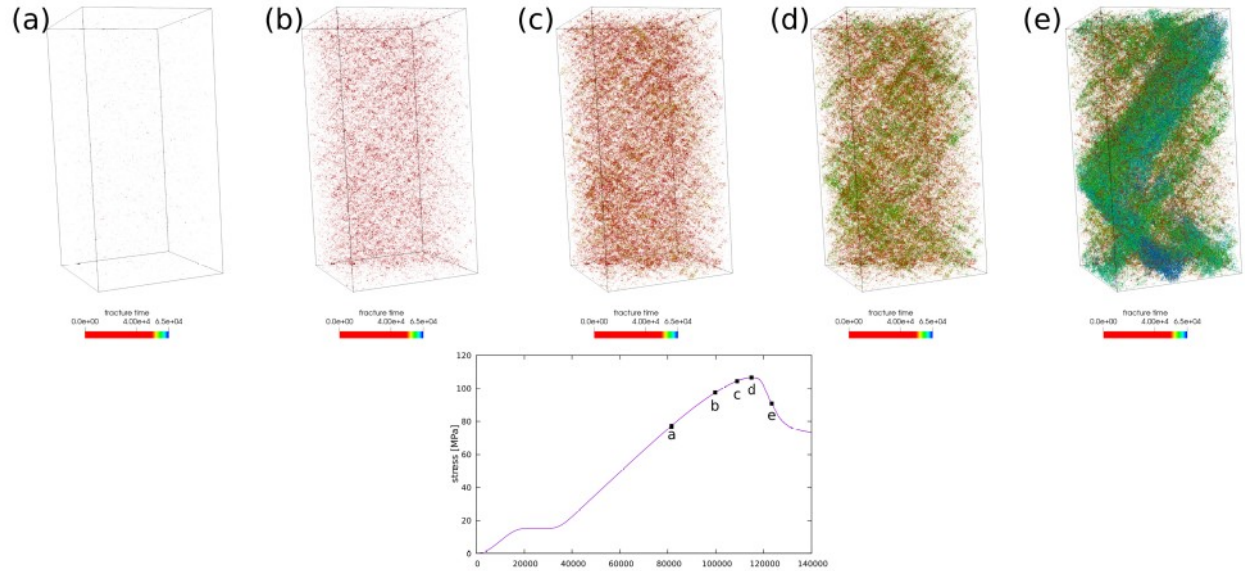

Figure 3: Evolution of micro-crack distribution in model with triaxial loading

The figures will be added to the supplementary material of the paper.