<u>Reply to Reviewers Comments</u>

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 1

This manuscript explores the possibility of using numerical modelling to hint the morphology of fractures by comparing them with results from laboratory experiments. The modeling is accomplished using DEM methods able to recreate a reliable shape of fracture surfaces, then compared to natural fracture surface from lab experiments. The Authors claim to have explored this comparison in the various stress conditions, yet due to the small size of the generated natural fractures, this comparison is restricted to the two experiments developed in tri-axial conditions. This significantly reduces the impact of the manuscript, that may correctly focus on the used comparisons and requires a more focused title and abstract.

Even in the only two compared experiments there are substantial differences in the roughness results between numerical and experimental results. This mostly derives from the confusion between extensional fractures and faults (i.e. shear fractures). The experimental fractures show a clear shear-related formation (their reported angle to the stress conditions is exactly what expected for faulting).

The fact that the experimental fractures are formed under mode 2 (shear) conditions is the consequence of the applied stress conditions. However, this applies also to the majority of the numerical models, except for one set of unconfined direct tension tests as described in section 3.1., lines 163-167. As a result the angle of the fractures relative to the stress orientations is very similar between the numerical fractures (Fig 1b, Fig 4a) and the experimental ones (Fig 4b,c).

And this might well be the main factor that produces the found roughness differences!

We did consider the possibility that the difference in the observed roughness between the numerical and the experimental samples is due to differences in the conditions under which the fractures formed. However, we do not believe that those differences are caused by additional shear motion on the fault in the laboratory experiments compared to the numerical models for several reasons:
First is that the available literature suggests that the small shear displacement occurring in the triax test has only a marginal effect on the roughness, i.e the change in the Hurst exponent H is less than 0.1 [Armitrano & Schmittbuhl 2002, Fig. 12], and that at low confinement H actually goes <u>down</u> with shear. This suggests that the Hurst exponent determined for the lab samples may differ from the value which would have been obtained at zero shear offset some very small amount, but that is clearly not sufficient to explain the difference between lab and numerical experiments.
A second reason is that, on close inspection, the grooves in the limestone sample (Fig 4b) are clearly not caused by abrasion due to the shear offset which might have occurred after fracture. In particular the size of the grooves (several mm deep, ~1cm wide) is too large to be caused by the slip which might have occurred on the surface after the initial failure, which is, specifically under our experimental conditions, at most a few mm (less than 2mm).

The third reason is that we did run a small set of simulations which we did not stop immediately after the formation of a through-going fracture be where we instead continued to deformation until an axial strain of ~12% was achieved and took snapshots of the fracture surfaces at regular data. While it is not clear how much shear offset the experimental fracture surfaces actually did experience, it was clearly much less than the ~13mm of axial displacement which, given the model length of 110mm, would be equivalent to the 12% axial strain in the numerical models.

The obtained Hurst exponents did show no significant trend with increasing strain of the model and offset of the shear fracture. Of the 3 models, and therefore 6 fracture surfaces, 3 showed no significant change of the Hurst exponent at all, i.e. H stayed in a range of +- 0.02 of its initial value (038L H2, 040L H2 and 043L H1 in the Figure 1 below), two showed an increase of H between 0.05 and 0.1 (040L H1 and 043L H2 in Fig. 1) and one (038L H1) even shows an initial drop in H for small strains, which is not fully recovered by the subsequent increase at larger strains.
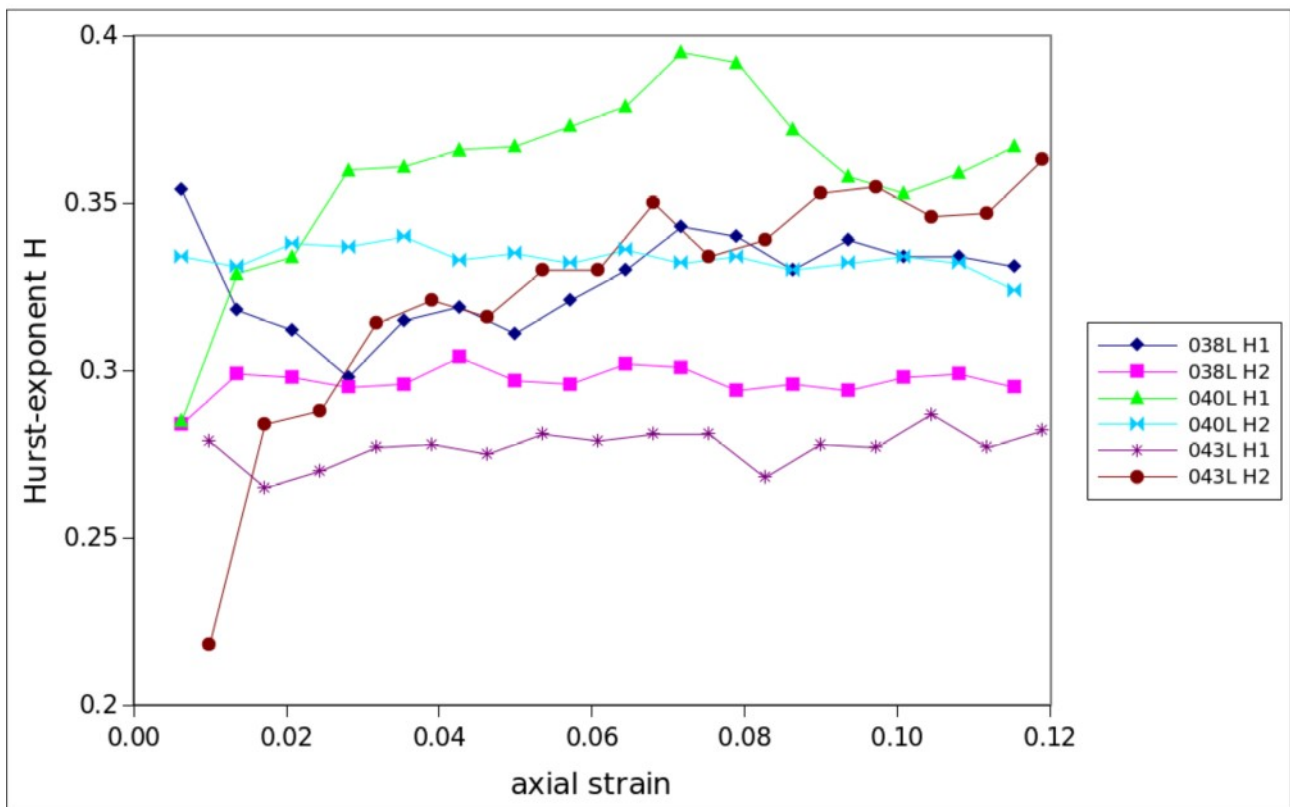

Figure 1: Hurst exponent of fracture surfaces vs. axial strain of deformed sample for 6 surfaces from 3 numerical models. In the legend, the 1st part of the description (038L) is the model ID, the 2nd part (H1, H2) distinguished between the two surfaces of the fracture.

While the average of the Hurst exponents from the 6 surfaces (Fig. 2 below) could be considered as showing a slight increasing trend for strains up to ~8%, the increase of 0.03 is about an order of magnitude too small to explain the observed differences between numerical and experimental surfaces. However, it would be compatible with the effect observed by Armitrano & Schmittbuhl [2002].

We did not include those data in the manuscript for two reasons, one being that a systematic study of the relation between shear offset and roughness parameters of fault surfaces was not within the scope of the work, the other being that of course the data shown above are not sufficient to draw any statistically valid conclusions on the details how the Hurst-exponent evolves with shear. However, given that the data does provide a strong suggestion that the observed differences in the Hurst-exponent between numerical

and experimental fractures are not due to possible differences in the amount of shear, they can probably be included in the discussion section of the manuscript. We will therefore add a paragraph discussing this issue to the manuscript and the Figures 1 and 2, and also
80   Figure 1 from the reply to reviewer 2 (page 5 there) to the supplementary material.
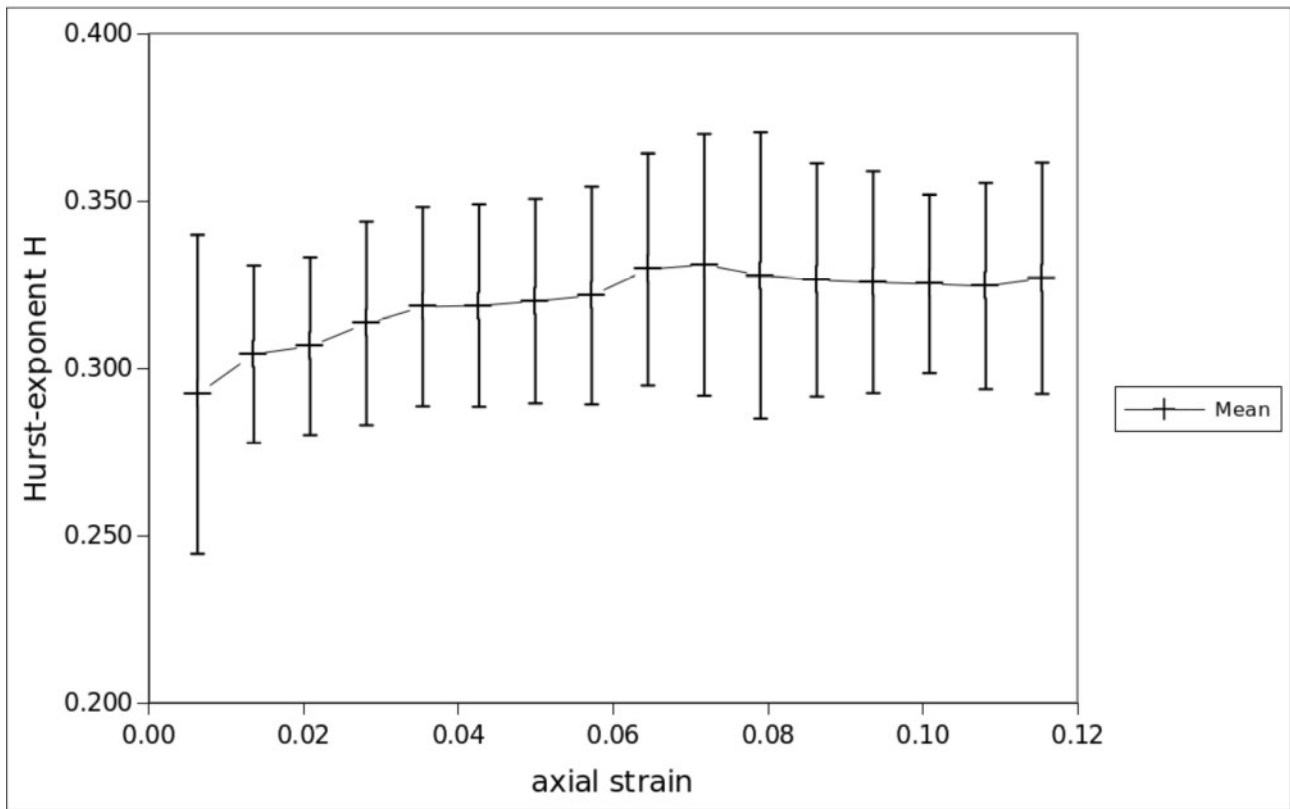

Figure 2: Average of the Hurst exponents of the 6 surfaces shown in Fig. 1 vs. model strain. Error bars show standard deviation.

The slip along the surface, even if very limited, is responsible for a smoothing process (that, in the case of large faults would derive in to the well-known "fault mirrors",
85   *characterized by the almost absence of roughness along the slip direction).*

We agree with the reviewer that slip along the surface can result in a significant reduction of the Hurst-exponent for profiles parallel to the slip direction down to values below 0.5 as, for example, shown in Table 1b in [Candela et al., 2012]. However, those large reductions
90   appear to apply mainly to faults with large amounts of slip, i.e. several meters up to kilometers. Data available in the Literature [Amitrano & Schmittbuhl 2002, Davidesko et al., 2014, Badt et al. 2016] suggests that at least under brittle conditions and for shear amounts less that a few cm the Hurst-exponents remain largely unchanged. As mentioned above, shortening in our experimental tests did not exceed 2mm.
95

This smoothing is proportional to the slip amount, yet could be easily replicated in the numerical modeling (that are intentionally by the author halted at the rupture point, even if the slip starts immediately DURING the enucleation of the shear-fracture surface). This additional behavior can be accomplished by adding the proper directional smoothing
100   *kernel on the re-oriented rupture plane (on the "height" component?). My suspect is that the proper application of this filter would significantly reduce the difference in the roughness between numerical modelling and experimental results.*

It appears that, in particular for small amounts of shear, the statement that "*smoothing is*
proportional to the slip amount" may not be generally applicable in all situations. It has in
fact been suggested in the literature that the roughness evolution of fault surfaces might
actually undergo a roughening stage [Davidesko et al., 2014, section 5.3]. If this is indeed
the case, applying an appropriate filter to the surface data to simulate the effect of shear-
induced smoothing as a post-processing step would be highly non-trivial.

On the other hand, more information are required on the grain-size of the DEM modelling.
We know that the efficiency of a stress to break grains is inversely proportional to the grain
dimension. That is, the grain-size distribution used in the DEM may affect the roughness
of the results. My suggestion to the authors is to include this factor in the modelling and to
explore its role on the roughness in the numerical experiments.

We agree with the reviewer that size dependent comminution of grains can clearly have an
influence on the evolution of the roughness of fracture surfaces. However, to include this
effect into the numerical models would increase the computational cost by at least an
order of magnitude if using an empirical model to break individual particles and replace
them with a set of smaller particles like the one proposed by Cleary [2001] and probably
more if the grain fracture process is actually modeled fully as in Abe & Mair [2005]. An
inclusion of grain fracture into the models would therefore result in impractically long
computing times, even when using current high performance computing facilities. This
would apply in particular to the simulation of statistically useful model ensembles. We
therefore did not include intra-grain fracture into the numerical models and we explicitly
mention this as one of the differences between the models and real rock (Discussion, lines
365-374, Conclusions, last sentence).

On last point is about the discarded fractures related to their small dimensions… I did not
fully understand it. Was it a problem of resolution in the scanning with respect to the
particle dimensions of the DEM (i.e. to have comparable number of points)?

In our samples we had the difficulty, that the deformed samples disintegrated into small
pieces. However, there is a limitation in the size of samples (=planes) which can be
analysed. The key problem is that with the available tools for the analysis of the
experimental fracture surfaces, i.e. photogrammetry using a standard digital SLR with a
macro lens (section 3.2, lines 223-227), the resolution of the obtained surface data is
restricted to about 0.1mm. Given that the analysis of the surfaces roughness required at
least around 100 sample points per direction (section 4.2, lines 327-330), and the fact that
the sampling interval should be somewhat larger than the resolution of the data, this
restricts the minimum size the usable surfaces to at least 2-3cm in each dimension of the
surface. Unfortunately such large fragments were produced only in very few of the
deformation experiments. The photo below shows the typical remains from one of the
triaxial experiments.

Figure 3: Fragments of a limestone sample after a triaxial deformation experiment.

What about improving the scanning resolution or enlarging the samples in the experiments?

While this would theoretically be a possible way to circumvent the problem, it is unfortunately not feasible within the current study due to the lack of the necessary tools and resources to do so. In particular, the dimension of the rock samples (55mm diameter x 110mm length) had to be chosen as standard sample size to fit the available deformation apparatus.

It would be of great interest to fully explore the generation of real, small extensional fractures.

We agree that this would be an interesting topic. However, to do this one would need different techniques for acquiring the necessary height field at the required resolution. We also must stress that the investigation of real, small extensional fractures was not the topic of our paper.

Here is a list of observations on the text:
Line 6 and forward- why using the term self-affine with respect to self-similar (e.g.
170 *Turcotte, 1992, areference that is strangely missing)*

We are using the definition from Mandelbrot [1985] and Bouchaud [1997] defining self-similarity as statistical invariance under isotropic scaling and self-affinity as statistical invariance under an affine transformation. To quote Bouchaud [1997]: "Unlike self-similar
175 objects, self-affine structures, being intrisically anisotropic, are not statistically invariant through a global dilation but rather through an affine transformation". This should, to our understanding, actually be equivalent to the description of a self-affine fractal on page 8 of [Turcotte 1992]. In regard to the description of rough surfaces Turcotte [1992] actually states that "A cross-section of topography with elevation plotted against position along a
180 linear track is not a self-similar fractal; however, it is usually a self-affine fractal."
The reference to [Turcotte 1992] can of course be added to the relevant section of the manuscript.

185 *Line 30 it is better to use resolution rather than "scale"*

In the context of that sentence the term "scale" is actually the correct one. The RMS deviation of a surface from a given average plane depends on what scale, for example mm, meters or kilometers, it is measured, not so much at what resolution that deviation is
190 sampled.

Line 53 Please indicate explicitly the dimensions of the model (3D re-projected on a 2D+height dimension?)

195 As there is no "model" mentioned in line 53 we are not completely sure what the reviewer meant here. This section (2.1 Discrete Element Method) only describes the numerical simulation approach as such, not the specific models used with that approach later in the manuscript.

200 *Line 68 cylindrical since is a symmetrical 2D model? This is not clear*

The assumption of cylindrical bonds between individual DEM particles is a fairly common way to parameterize the brittle-elastic interactions between those particles [Potyondy & Cundall 2004, Weatherly 2011]. The choice of a circular cross section for the bonds is
205 largely due to the fact that the particles are assumed to be spherical, but it is also mathematically convenient.

Line 87 why 30 degrees? Is it related to the frictional angle, as suggested in: (e.g.) A Nur, H Ron, O Scotti,1986?
210
No. This parameter has no physical meaning at all. When developing and testing the algorithm, the 30 degrees separation between the different view directions used to identify surface particles has been found to provide a good balance between computational effort and quality of the reconstructed surfaces. Figure 2 in the manuscript will be changed to
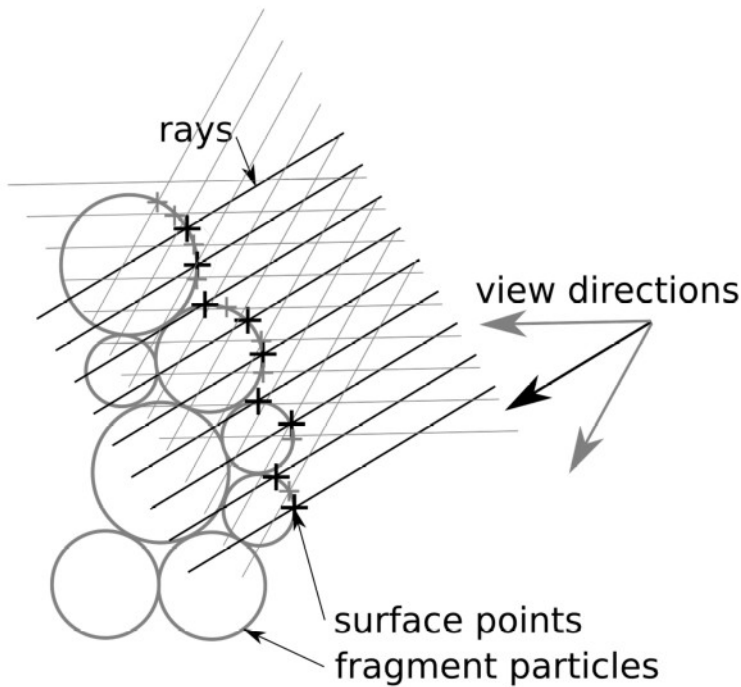215 make this more clear. See modified figure below.

Figure 2 (modified): Simplified 2D sketch of the ray-casting method. The gray particles are assumed to belong to the same fragment of the deformed sample and the black crosses show the fragment surface calculated from the line-particle intersections. **Black lines and black arrow show primary view direction, light gray lines and arrows show additional view directions at a 30 degree angle to the primary direction.**

Line 90 more info on this "final step" will ease the understanding of the numerical process

220

The final step to determine which other fragment of the model is closest to a given surface particle essentially involves checking all other particles in order of their distance to this particle if they belong the same fragment of the model as the surface particle or to another large fragment of the model until one is found belonging to another fragment. On a more
225 technical level it does involve the use of an acceleration grid to avoid the calculation of all particle-particle distances while still being able to pre-sort the particles roughly in order of their distance to the surface particle for which the calculation is performed.

Line 106, 126 and in the manuscript: there is some confusion among the terms "height",
230 *"y'", and "z'". It seems they might indicate the same coordinate axis. Please explicit or indicate the differences*

"Height" generally means a distance from a given plane measured along the plane normal as described in line 131 and Fig. 3A in the manuscript.
235

Line 153 please add indications on how the random selection is accomplished

The software (python script) written by the authors for the surface analysis uses the function "numpy.random.choice()" from the numpy-library (www.numpy.org) to draw a
240 random selection from an array of points. Technical details of that function can be found in the online documentation of the numpy-library (https://numpy.org/doc/).

Line 157: it would be useful if you indicate here the relation between the Hurst exponent and the fractal dimension D

The Hurst H exponent and the fractal dimension D of an object are related as D=2-H for a 1D-profile [Mandelbrot 1985] or, more generally, D=n+1-H where n is the dimension of the object [Yang & Lo, 1997], i.e. n=1 for a profile and n=2 for a surface. We will add this explanation in the modified manuscript.

Line 163. Here is where a full description of the particle size distribution in the numerical experiment should be added

The insertion based particle packing algorithm by Place & Mora [2001] mentioned in line 194 does produce particles with a power-law size distribution with an exponent of approximately -3. I.e. the number of particles in a given size bin is proportional to r^{-3}. See Figure 3 below for an example.
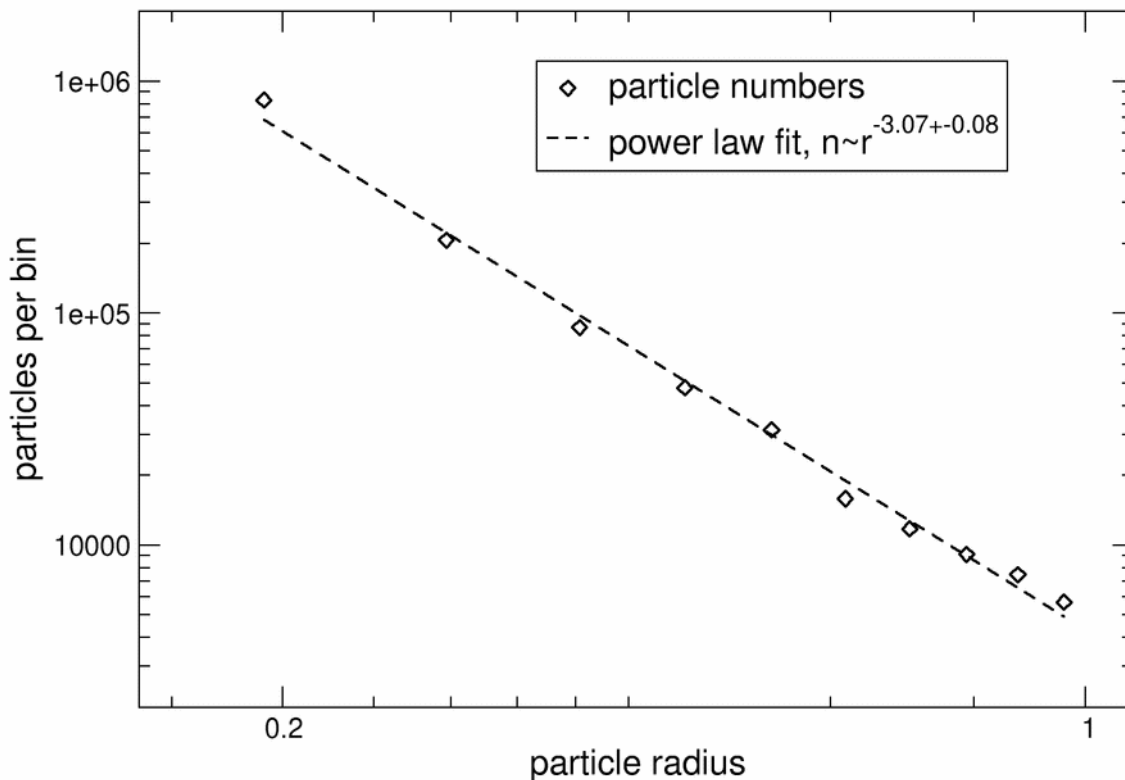
Figure 4: Log-log plot of the size distribution of the particles in one of the models.

Line 163 : "a large number"
Line 165 : s1 > s2 > s3 >0

Will be changed.

Line 179: please indicate which is the "chosen speed" and why it was chosen since this will affect the rupture roughness

As described in lines 195ff, the loading speed was set to 17cm/s in order to be sufficiently similar to experimental values while keeping the required computational cost at a viable level. Given that this speed is nearly 4 order of magnitude below the dynamic fracture propagation speed, it should not influence the processes happening during rupture propagation to any significant degree.

Line 186: please quantitatively specify "rest" and indicate how it has been chosen

We are assuming this concerns line 184, not 186.
The "rest" phase was chosen to be 10000 time steps in the numerical model. This was sufficient for the particle kinetic energy to be reduced by about an order of magnitude.

Line 195-198 the quantitative comparison between experiments at different speed would prove this sentence

Due to the sensitivity of the fracturing process to small perturbations, the speed at which a model is deformed can indeed influence the outcome of individual simulations, making the direct comparison of experiments at difference speeds difficult. In order to compare the outcomes of such experiments statistically we therefore performed a set of 6 simulations at a reduced compression speed of ~8.5cm/s, i.e. half of the rate used in the majority of the simulations. All those simulations were performed under transverse isotropic stress conditions, i.e. $\sigma_2=\sigma_3$. The analysis of the 20 fracture surfaces produced in those simulations showed that neither the Hurst-exponent nor the RMS roughness or Joint Roughness Coefficient (JRC) were statistically significantly different from those of surfaces produced with the faster compression rate under the same stress conditions. The average Hurst exponents obtained from the "slow" experiments was H=0.408 with a standard deviation of 0.064 whereas for the "fast" experiments it was 0.375±0.084. The RMS roughness was 2.23±0.45 for the slow vs. 2.39±0.77 for the fast models. The calculated JRC is 28.9±1.9 for the slow and 29.7±2.8 for the fast models.

Line 202 among

Line 218-221 Please rephrase Abstract and Introduction accordingly!

We are not sure which of the aspects mentioned in those lines the reviewer refers to. We could explicitly mention in the abstract and the introduction that we only had a small number of usable samples from the laboratory experiments available.

Line 226: Apologize for my ignorance: what is mio unit?

Millions. This will be changed in a revised manuscript.

Line 224-227 To complete the indications, the distance between lens and object must be specify, since this influence the resolution

The distance between object and lens was variable, but usually slightly larger than the minimum focus distance of the lens, which is approximately 5cm, measured from the front lens to the object or, according to lens specifications, 16.3cm if measured from the camera

sensor. I.e. the photos were taken at varying distances between 5cm and 10cm from the front lens to the object.

320 *Line 240 were calculated, were*

Will be changed.

Line 264-267 as discussed earlier, the found angle is obvious: faulting (shear plane)

325

The angle was only mentioned here to highlight the difference in orientation, and therefore maximum possible size, between the fractures generated by shear failure due to triaxial compression of the model and those generated by tensile failure due to uniaxial extension of the model.

330

Fig.9 indicate the unit of the x-axis "Distance" scale and its relation to particle size

The unit used for the distance scale is the radius of the largest particles in the DEM model.

335 *Line 281-287 as discussed earlier, this is something that has to be fully discussed earlier for its general meaning*

We probably should made this statement more precise to avoid a possible misunderstanding. The sentence "Only the absolute value of the roughness depends
340 somewhat on the size range of the particles." only applies to the "cut" surfaces, not to surfaces obtained by numerical simulation of fracture.
We will change this sentence to "Only the absolute value of the roughness **of the cut surfaces** depends somewhat on the size range of the particles.".

345 *Line 304-307 a figure is require to prove it*

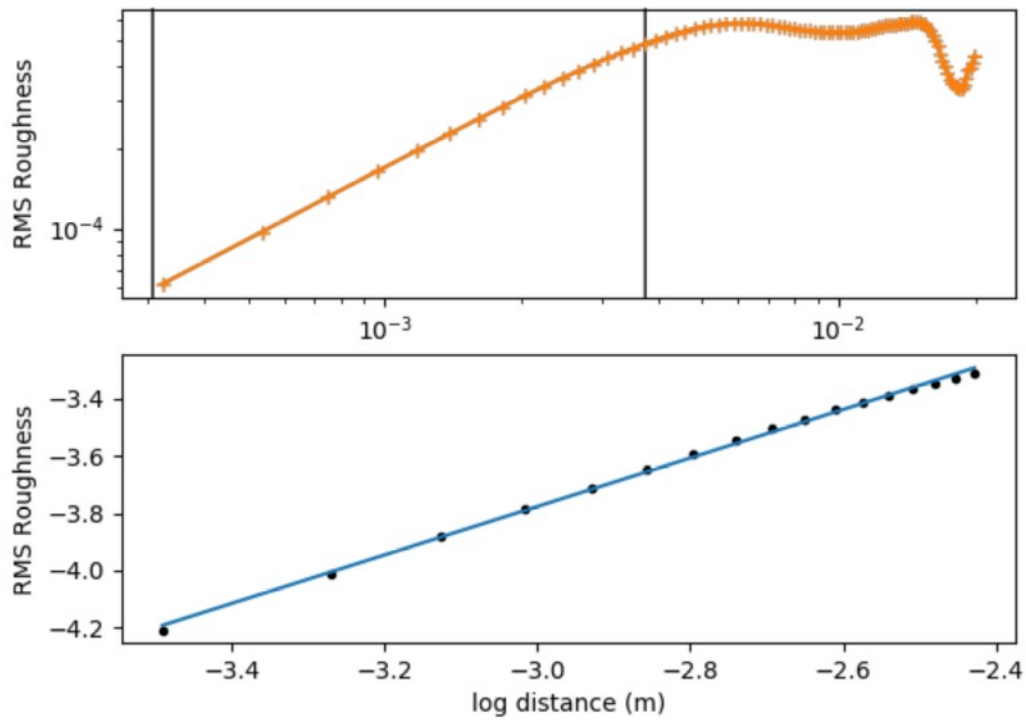See Figure 5 (Limestone sample) and Figure 6 (Sandstone sample) below.

Figure 5: Height-height correlation function for the Limestone sample. Top: Full distance range, bottom: range used for fitting a linear relation and calculation of Hurst-exponent.
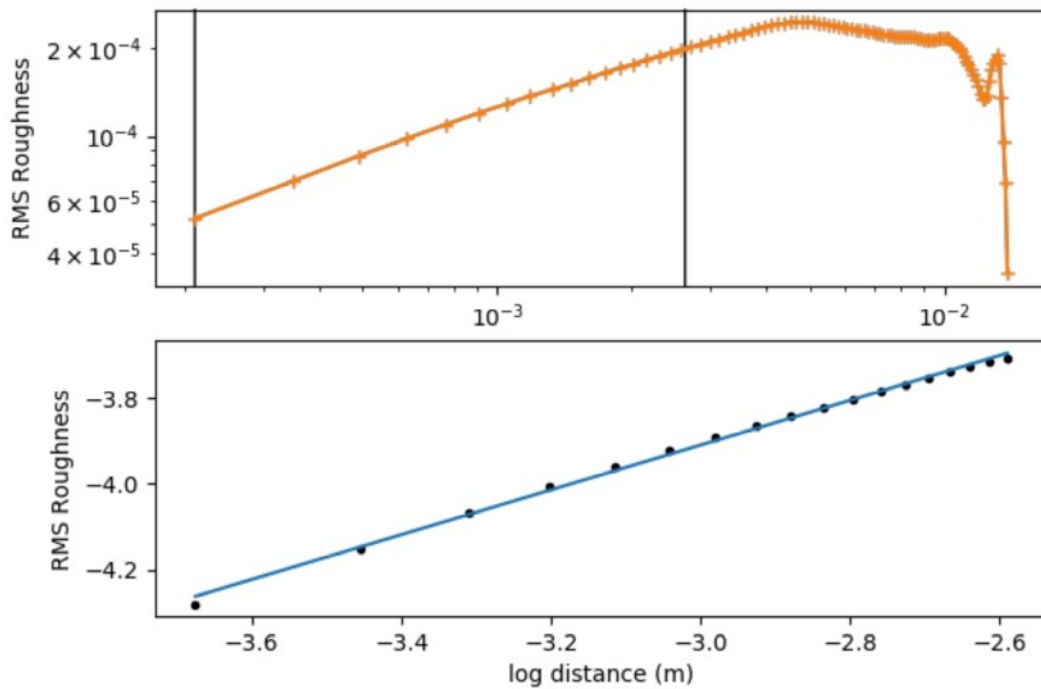

Figure 6: Height-height correlation function for the Sandstone sample. Top: Full distance range, bottom: range used for fitting a linear relation and calculation of Hurst-exponent.

Line 315-317 is this statistically meaningful?

355 As there are only 2 data points where this effect is observed, the statistical significance isn't really clear. The statistical argument that those two points are indeed outliers would be their H-value is more than 2 standard deviations below the closest other value (0.5 vs. 0.67, standard deviation is 0.08). The other argument would be that they are the only data points where we don't see a clear single linear section in the height-height correlation 360 function.

Line 318-322 Does this mean that their distribution is not self-similar ("self-affine")? A comment will help

365 On the contrary. As stated in lines 322/323, "It is also to be expected based on the fact that the analyzed surfaces are self-affine.", i.e. the observations described in lines 318-322 are in fact a consequence of the self-affine properties of the surfaces.

Line 339-344 refer to what presented earlier

370
We do not fully understand this comment. We refer to what was already presented in Line 341: "The results of the analysis of the simulation data (Section 4.1) shows that..."

Line 347-356 According to this sentence, the manuscript should be reshaped to avoid
375 *creating disconfirmed expectations by the readers*

We are not sure what the reviewer means with "disconfirmed expectations".

Line 357-360 the use of two different measure (% and ration) might be confusing for the
380 *readers*

We would prefer not to change this sentence.

Line 364-380 refer to what discussed earlier

385
As discussed in lines 120-131 of this reply, the inclusion of a grain size reduction mechanism would be interesting for further study, but is currently outside the scope of this work.

390 *Line 367-369 add a comment on grinding and smoothing produced by sliding*

We will add the following sentences: "Smoothing due to abrasion while sliding is, in general, an important mechanism for the modification of rough surfaces. However, experimental data published in the literature [Amitrano & Schmittbuhl 2002, Davidesko et 395 al., 2014, Badt et al. 2016] suggest that it is unlikely to have a sufficiently large effect at the small shear offsets in both numerical models and experimental samples studied here to explain the observed differences. "

Line 403 apologize again for my ignorance: what is "Brasilian test"? Maybe some
400 *description of it for "the rest of us" will improve the reader comprehension*

A "Brazilian test", also known as "Brazilian disk test" is a standard geomechanical test to obtain the tensile strength of materials by applying a compressive load to two opposite points on the circumference of a circular disk of the material. This loading results in a

405   tensile stress in the central part of the disk which is oriented perpendicular to the line connecting the two loading points. It is often preferred to a direct tension test because it is easier to perform. However, it only works if the material has a sufficiently high ratio between compressive and tensile strength, otherwise it will fail in shear instead of tension. See, for example Jaeger et al., [2007] or Fjaer et al. [2008].

410   An appropriate reference will be added.

Line 409-411 I strongly doubt this, according to the relation with the role played by grain-size distribution (see earlier). Is this (forced) sentence necessary? Or better justify it!

415   In order to clarify that this sentence is meant to be understood in the context of the aforementioned observations by Nigon et al. [2017], we will modify it to "This difference in scales shows that the Hurst-exponents in our numerical models are completely calculated above the "transition scale" of Nigon et al. [2017] and therefore should belong to the regime described as "jointing induced roughness" by them. This means that the low values

420   of the Hurst-exponents in the numerical can not be explained by the "grain induced roughness" regime of Nigon et al. [2017]."

Line 415-416 at least the values of the results must be mentioned. As it is written, it seems a subjective elimination of undesired data…

425

The relation between Hurst-exponents and JRC for the limestone sample could, in principle, be analyzed separately for the directions parallel and perpendicular to the shortening direction to take the strong anisotropy of the JRC into account. If this was done, the data for the shortening-parallel direction (H~0.55, i.e. D-1=0.45, JRC~6.5...7.5) would

430   plot in Fig. 17 even further below the relation proposed by Ficker than the sandstone, whereas the data in the perpendicular direction (H~0.605, D-1~0.395, JRC~16...17) would be closer, but still below.

Line 422-435 again the problematics connected to grain-size distribution. My suggestion is
435   *to discuss this earlier in the text (in the introduction?)*

The data (Fig. 14 in the manuscript) suggests that the variability of the Hurst-exponents is likely be influenced by the "resolution", i.e. the ratio between grain size and surface extension. However, the mean value of the Hurst-exponents doesn't seem to be

440   influenced. In addition, a small number of tests using numerical models with a different particle size distribution, specifically a larger range of particle radii has also shown no influence of the particle size distribution on the Hurst-exponent in these models. See also our response to Reviewer 2, lines 105ff.

It is of course possible that the grain size distribution might have an effect if grain fracture

445   is included in the model. However, as explained in lines 368-372 in the manuscript, this would be beyond the scope of the current work.

Line 429 The experiment velocity (i.e. velocity of propagation of fracture surfaces) will influence the re-organization of stress and the influence of local secondary stress
450   *components, and therefore the resulting roughness. This has to be taken into account when presenting results. Perhaps a series of experiments at different velocity will highlight this very interesting point.*

Not shure how this relates to line 429, but care should be taken not to confuse the <u>loading</u>
455   <u>velocity</u>, i.e. the velocity with which the deformation is applied to the top and bottom surface of the sample in the numerical experiments, with the <u>fracture propagation velocity.</u>

The first on is a boundary condition, which can be controlled by changing a parameter in the simulation, and is set to ~17cm/s in the simulations presented here. The latter is a result of the internal dynamics of the simulated material and is purely controlled by its elastic parameters, i.e. shear and compressive modulus and density.

It would of course be possible to change the fracture propagation velocity in the models by varying the elastic properties of the simulated material, but given that the fracture roughness data published in the literature seem to be largely independent of rock type, and therefore rock elastic properties, it is not clear whether this would actually be a promising direction.

Line 443 please specify that they result from tri-axial conditions! (and therefore the development of faults -shear surfaces)

We will change "...Photogrammetric analysis of fracture surfaces" to "...Photogrammetric analysis of shear fracture surfaces"

Line 451 better "result" rather than "phenomenon" ?

Will be changed.

References:

Badt, N., Hatzor, Y. H., Toussaint, R. and A. Sagy, 2016, Geometrical evolution of interlocked rough slip surfaces: The role of normal stress, Earth and Planetary Science Letters 443, 153–161

Cleary, P.W., 2001. Recent advances in DEM modelling of tumbling mills. Miner. Eng. 14, 1295–1319

Davidesko, G., Sagy, A. and Y. H. Hatzor, 2014, Evolution of slip surface roughness through shear, Geophys. Res. Lett., 41, 1492–1498, doi:10.1002/2013GL058913

Fjaer, E., Holt, R.M., Horsrud, P., Raaen, A.M. and R. Risnes, 2008, Petroleum related Rock mechanics, 2nd edition, Elsevier

Jaeger, J.C. Cook, N.G.W. and R.W. Zimmerman, 2007, Fundamentals of rock mechanics, 4th edition, Blackwell

Potyondy, D.O. and P.A. Cundall, 2004, A bonded-particle model for rock, Int. J. Rock Mech. Min. Sci., 41, 1329-1364

Turcotte, D.L., Fractals, chaos, self-organized citicality and tectonics, 1992, Terra Nova, 4, 4-12

Yang, Z.Y. and S.C. Lo, 1997, An Index for Describing the Anisotropy of Joint Surfaces, Int. J. Rock Mech. Min. Sci., 34(6), 1031-1044