Below, questions and suggestions are presented concerning particular passages in the text (referenced by the line numbers in the manuscript):

line 83:

Replacing „The deformation process is necessary for the long term safety case 84 analysis for HLRW repositories“
by “A sufficient understanding of the deformation process is necessary for the long term safety case 84 analysis for HLRW repositories“
would be more precise.

line 98-100:

The importance of sealing for the exclusion of oxygen from the sample to prevent oxidation of pyrite and subsequent formation of gypsum should also be mentioned.

line 109:

The reference for the picture is missing.

line 156-170:

The information concerning the resolution of the employed CT scanners gives a good idea of their capabilities and limitations, but it is not consistent in detail:
The speed|scan CT 64 is characterized by “the spatial resolution with typical 0.5 to 1 mm”, whereas the resolution of the performed scan (312 µm) is outside this range.
Consequently, the resolution of the v|tome|x L300 with “approximately 60 µm” is only 5 times better than that of the speed|scan CT 64 – not “a factor of 10 better” as stated in line 166.

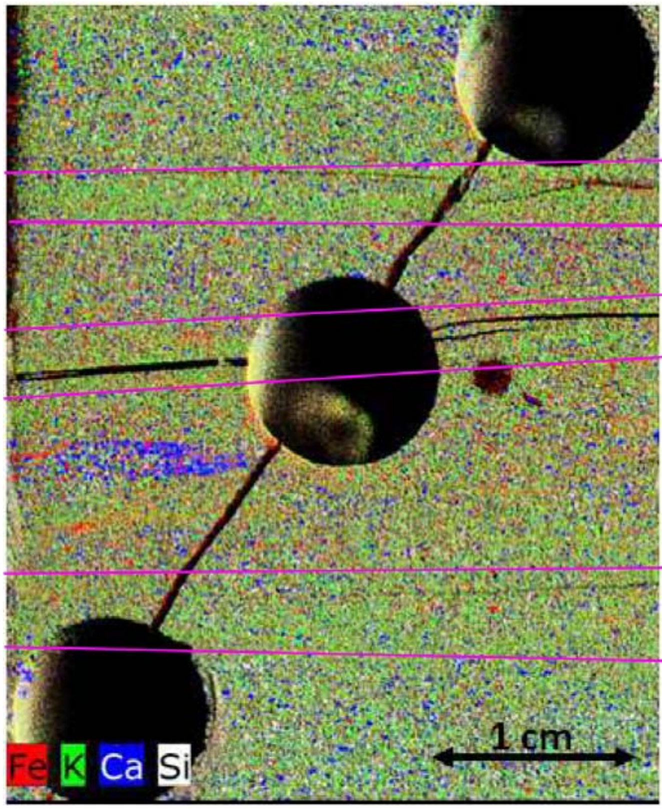
line 212:

Replace “were” with “where”.

line 208-215:

The given description of Figure 4 (particularly its left part) is only partly comprehensible to me:

“In these small sections of about 5 mm, bedding features could not be detected anymore.” –
This is true for the 5 mm sections. But it does not apply to the left part of Figure 4 (showing a pixel size ≤2 times the pixel size in the 5 mm sections). Indeed, on this scale bedding features are no longer dominating the optical impression, but they are still easily detectable. Several layers clearly associated with the orientation of the bedding are shown in an annotated copy of Figure 4. This is worth mentioning because it can provide an indication of principles of crack generation or propagation in this material (see below).



"Instead a few 50-100 µm thick bands of either carbonates (blue) or clays (green) could be observed with a significant angle compared to bedding." – Locking at the right part of Figure 4, I cannot really identify these bands with a significant inclination with respect to the bedding. If they are there, think on adding some annotation to the figure to assist the observer.

"Assuming that these small lineaments were no XRF artefacts, it can be supposed that a crack started to form there. … Therefore it can only be assumed that the small lineaments observed both with the XRF scanner as well as with the SEM could be the starting point for crack formation." – this appears to be too speculative because there is no evidence available for this assumption. In fact, there is some evidence that crack formation might coincide with material heterogeneities related to bedding features: There are 3 cracks (resp. 3 systems of cracks) in bedding orientation (disking cracks) present in Figure 4; each of them is developed within the layers that exhibit the highest clay content (cf. the annotated copy of Figure 4). Hence, the relative weakness of clay rich layers obviously is relevant for the generation and propagation of cracks in the investigated claystone.
Keep in mind that the 5 mm sections are just 2D images with a rather limited number of pixels. Therefore, some smaller "lineaments" necessarily appear simply by random distribution of mineral particles. Furthermore, there is no evidence that these "lineaments" are structures with a 2-dimensional extension which would be an important requirement for their mechanical relevance. As far as I can see, the only extended heterogeneity visible in the 5 mm sections, which is clearly related to the fractures appears in the first section: The crack separates a clay rich area (left) from a region with high content of pyrite (right). Hence, this boundary between regions of different composition predetermines the propagation of the crack, but there is no evidence that crack formation did start at this feature.

"The location from ~~were~~ where tension relief observable as crack formation started is assumed to be outside the investigated area." – Which cracks does this statement refer to?

- There is one crack orientated diagonally in Figure 4. The formation of this crack most likely started outside the investigated area. But this crack definitely did not originate as a tensile crack (and therefore has nothing to do with a "tension relief"). It is a typical shear failure plane that develops in a triaxial strength test under sufficient confining pressure. The opening of this type of failure plane does not happen until the overall compressive stress state during the strength test is terminated.
- There are 3 cracks (resp. systems of cracks) in bedding orientation present in Figure 4. These cracks are most likely formed by tensile failure. As far as I can see, there is no evidence that the formation of these cracks started outside the investigated area. In contrast, the observation that these cracks split into systems of interconnected subparallel cracks in the vicinity of the shear failure plane (cf. Figure 4) might offer some indication that their formation started at the shear failure plane. Regarding the small scale heterogeneity of the stress field occurring along the shear failure plane during shearing (due to inhomogeneous friction as well as to unevenness of the shear failure plane), an initiation of tensile cracks in this area seems likely.

Thus, whatever crack might be addressed in line 212/213, to me the statement appears to be unconvincing (for different reasons depending on the addressed crack).

line 216:

The last 2 pictures in the right side column have to change places to correspond to the pictures in the central column.

line 227-229:

"Interestingly, the shear fracture is located within the clay rich area of the OPA sample. Starting point is right at the border between clay carbonatic zone (right hand side of Figure 5)." – This statement, in particular the second sentence, cannot be proven by a 2D-picture as it is presented by the virtual slice in Figure 5 (the intersect of the failure plane with the cylinder barrel could extend considerably into the carbonatic zone in parts not visible in Figure 5). But checking "Video 3 - speed scan data set.avi" in the supplement proves that the statement is correct. Thus, adding a reference to the video in this context might be a good idea.

line 253:

"Besides" instead of "Despite" appears more appropriate.

line 261:

Replace "carbon shells" with "carbonate shells".

line 262-264:

"Accordingly, smaller shear fractures can be detected, which are more or less parallel oriented to the main shear crack (Figure 7, left hand side)." – There are at least 3 types of cracks visible in Figure 7:

- The long diagonal cracks which sometimes split up into several subparallel braches. These are the shear fractures constituting the shear failure plane.
- Some cracks following bedding plane features. Most likely, these have been formed by tensile failure.
- A system of small stacked cracks orientated more or less perpendicular to the large shear fractures. They have obviously formed later than the large shear fractures as they are truncated by the shear fractures. Generally, shearing normal to the main shear failure plane is expected to be almost negligible. Therefore, these stacked cracks are very likely generated by tensile failure.

This might be explained as result of a slight bending, that can affect the thin block of material between 2 large parallel shear fractures during the shearing.

line 298-301 and Table 1:

Generally, the conclusion of this statement is comprehensible and plausible. However, the methodic approach for the quantifying statements accompanying the argumentation is not at all trivial. Since the applied method for counting cracks and evaluating their frequency is not explained in the text, the basis of the argumentation remains nebulous and not well-defined. A number of questions arise regarding the determination of the “number of cracks” given in Table 1:
- Are these numbers derived from 2D-slices or from full 3D data?
- There is only one linear dimension (“core size”) given in Table 1. Hence it remains unclear, to which area (in case of 2D-slices) resp. volume (in case of full 3D analysis) the given crack numbers refer. Do all crack numbers refer to the same area resp. volume of the sample? This would restrict the analysis to a very small detail of the low-resolution scans. But it is mandatory to preserve comparability of the numbers, because otherwise scaling problems and effects of non-representative subsampling would make meaningful comparison of crack numbers almost impossible.
- Which criteria have been used to establish a well-defined method for counting cracks?

Basically, it has to be questioned whether the “number of cracks” is an appropriate measure to evaluate and compare the amount of information yielded by different CT techniques. First of all, a number of cracks has to be identified in a specified area resp. volume. Thus, a “crack number density” would be a more appropriate information. But there are still serious problems with this approach. Obviously, there is a scaling problem: Because cracks often range beyond the boundaries of an investigated volume, the number of cracks will not grow proportionally when increasing the investigated volume. Thus, the “crack number density” is a scale dependent quantity.

Another, even more severe problem arises from the fact that the determination of a “crack number density” inevitably requires the determination of a “number of cracks”. When using the term “number of cracks”, well-defined criteria are required for what has to be counted as “1 crack”. A consistent definition how to distinguish between “1 crack” and “2 or more associated cracks” is extremely difficult to achieve even in 2D. You can easily illustrate this problem by asking several people to determine the number of cracks visible in the left picture of Figure 8 – you will get a considerable variation between their answers. When switching to a 3D analysis this problem becomes even more severe.

Regarding these shortcomings, it is recommended to employ another measure to characterize the amount of information yielded by different CT techniques. This measure should exhibit considerable advantages compared to the “number of cracks” or the “crack number density”. In particular, it has to be determinable in a well-defined way, and it should be virtually unaffected by scaling effects. Whereas counting cracks turned out to be very problematic, the determination of crack area is a largely straightforward procedure offering the required advantages. Therefore, I recommend to use a “crack area density” (i. e. the area of detected cracks per investigated sample volume) instead of the “number of cracks”.

line 302-306:

It is evident that voxel resolution will have a considerable impact on determining the dimension of features that are close to or even below the voxel size. Nevertheless, the shown degree of this impact is surprising.

Partial volume effects are not sufficient to explain the amount of overestimation of crack aperture occurring in the large core scans: Even if voxels containing an almost negligible portion of crack are classified as part of the crack, the overestimation of the aperture w of a straight crack is limited to 2 times the voxel size: $w_{obs} \leq w_{true} + 2\, d_{voxel}$. On average, the impact of partial volume effects should be much smaller. Assuming that the average aperture determined on 3 mm core size represents the true value, the average aperture observed on 100 mm core exceeds this theoretical limit explainable by partial volume effects in any case (see table).

		average crack aperture [µm]			
		fracture A		fracture B	
core size [mm]	voxel resolution [µm]	observed	upper limit regarding partial volume effects	observed	upper limit regarding partial volume effects
100	312.5	990	807	1300	853
100	57.5	393	297	364	343
3	2.8	182		228	

Therefore, a significant part of the overestimation must be attributable to the impact of the effective segmentation resolution. Since this can vary significantly depending on the settings for numerous data processing parameters, adding another column to Table 1 showing the effective segmentation resolution is advisable.

line 310-312:

Identifying the fractures as “shear crack” and “disking crack” would be more informative than just numbering them “fracture A” and “fracture B”.

line 335-336:

“Nevertheless on the top of the shear zone a darker zone is identifiable, which is a result of particle reduction.” – Since X-ray attenuation does not depend on particle size (as long as particle size is large compared to the wavelength of the X-ray), but only depends on material density and the mineral composition, the darker zone should not be explained as “a result of particle reduction”. It is rather a result of loosening of the material (dilatancy) resulting in a lower density. Nevertheless, it is obvious from the overall impression of Figure 11-B and Figure 12 (and should hence be mentioned in the text) that this dilatant deformation of the material is accompanied by a reduction of particle size.

line 348-349:

“Close up SEM images (Figure 12) prove that the claystone did not simply break as one would expect from broken glass.” – This statement should be presented in a less generalized and more precise form.

The term “break” does not differentiate between shear failure and tensile failure, although these represent completely different failure mechanisms. As can be recognized from Figure 11, a mylonitic zone is only formed along shear cracks, whereas tensile cracks do not show any mylonitic features. Therefore, the statement in line 348-349 only applies to the breaking of claystone under deviatoric loading (shear). The statement is wrong with respect to tensile failure in claystone, which looks quite similar to the pattern one would expect from broken glass.

Due to the missing differentiation between different failure regimes, the given comparison between breaking claystone and breaking glass is misleading. One has to be aware that almost any case of breaking glass we know from our everyday experience represents a pure tensile failure mode. Thus, the difference in features observable on failure cracks, which is attributed to a material difference (claystone vs. glass) in the text, in fact is attributable to different failure modes (shear vs. tension).

line 354-355:

"It is not clear whether the mylonitic zone formed just before breaking or if it formed by the relative movement of both sides of the crack." – Taking into account the amount of local deformation and particle dislocation required to form the mylonitic zone, a formation before breaking would be very difficult to explain.