The study of Richter and Tsukamoto entitled "Investigation of quartz ESR residual signals in the last glacial and early Holocene fluvial deposits from the Lower Rhine" presents original results that are very important for the ESR dating community. The manuscript is well written. The main message is that Al-h as well a Ti-Li and Ti-H signals in quartz are not fully reset in the case of fluvial sediments. By analyzing Holocence samples that were securely dated by OSL for ESR dating the authors report a mean value of 1350 ± 120 Gy for the residual of Al centre, 610 ±60 Gy for the lithium-compensated Ti centre (Ti-Li), 170 ± 20 Gy for the hydrogen-compensated Ti centre (Ti-H), and 470 ±50 Gy for the signal originated from both the Ti-Li and Ti-H centres (termed Ti-mix), concluding that fluvial sediments carry a significant residual dose, and therefore the subtraction of residual dose using a modern analogue is highly recommended to obtain reliable ESR ages. I fully concur with the main message of the paper and I think it is worth being published.

We thank the reviewer for the very insightful comments to our manuscript.

However, besides this, the study presents yet another problem, namely the poor dose recovery results. The way the dose recovery experiment was conducted is problematic in my view and leads to circular reasoning. If the authors cannot prove that my observation is wrong, the interpretation provided now needs to be revised. The approach of the dose recovery test was to use the "natural" sample that holds this significant residual, irradiate it with 1000 Gy and as the authors explain "The dose recovery ratio was calculated by subtracting the natural De from the recovered dose and the difference of the natural + 1000 Gy and the natural De was then divided by the added dose of 1000 Gy." The problem is that by doing so, it is inherently assumed that the determined De is correct.

While this would not be a problem in a luminescence experiment with a residual of say 10 Gy compared to the given dose of 1000 Gy, here it is, because the magnitude of the residual, in this case the De is of the same order of magnitude as the given dose. The authors then state that "The dose recovery ratio for the Al signal is high with 1.74 ± 0.16, which indicates a sensitivity change due to thermal treatment during SAR protocol, therefore the reported residual doses may be overestimated." But to my understanding, first it was assumed accurate based on your experiment or am I missing something?

And you subtracted this value in the numerator of your ratio. The way this experiment stands defies in my view the idea of a dose recovery test as the given dose should be known and the equivalent dose is by definition an unknown. The only way in which this dose recovery could work is if the "sensitivity" is the same in the De (or residual dose here) measurement and during the dose recovery experiment. Which brings me to the point when I have to mention that I find the term "sensitivity change" to be extremely difficult to be conceptually integrated in ESR dating.

Anyway, conceptually the problem is that you draw a conclusion on the accuracy of the measurement protocol (namely that it is not very accurate) by feeding into your protocol a value measured by the same protocol that is assumed to be accurate.

Although we understand the criticism of the reviewer, implementing dose recovery has been something really missing in ESR dating to prove the robustness of measurement protocols.

The current version of dose recovery test is a simplified one of Tsukamoto et al. (2017), who conducted the single aliquot regenerative and added dose (SARA), after Mejdhal et al (1994) and Kars et al. (2014) using a single added dose point. Here, dose recovery ratio is the slope between SAR De and added dose. Assuming a same problem (if there is) is affecting the performance of the SAR protocol in both natural and natural + 1000Gy measurements, the slope should be equivalent to the dose recovery ratio. And since the full SARA measurement is very time consuming, the one point SARA as a dose recovery test is the best thing we can think of.

My second problem concerns figure 5 where the residual of Al-h is plotted as function of that of Ti-Li. While I fully agree with the logic of this approach and I consider it a very elegant way to extract the unbleachable part of the Al-h signal, even by eye one can see that that line does not fit the data. I plotted the data myself and obtained an intercept of 134 ± 208 Gy by simple linear fitting so I cannot understand how that regression was obtained. The result I have quoted above is of course very concerning, as I understand that the authors know from experience, by the simple examination of the signals measured on modern analogues from aeolian environments that Al-h is not fully bleachable. I fully support this view. I think that the fact that the intercept is close to zero and the scatter in your data simply reflects the problems associated with the measurement protocol, as mentioned in my comments above and below.

We admit that the linear trend line was drawn incorrectly. After reviewing the data, we agree with the trend line/intercept with the y-Axis of the reviewers comment. The figure was adjusted and the text was also adapted.

While one should not expect this problem to be solved in the framework of this particular study the observation "The De value of the Al centre was initially decreased by the preheat at 160 °C, but shows a steady increase in De with increasing preheat temperature. At 220 °C no De calculation was possible, because all regenerated signal intensities were below the natural" is extremely concerning. I suggest the authors show some of the measured signals and the constructed dose response curves in the supplementary file.
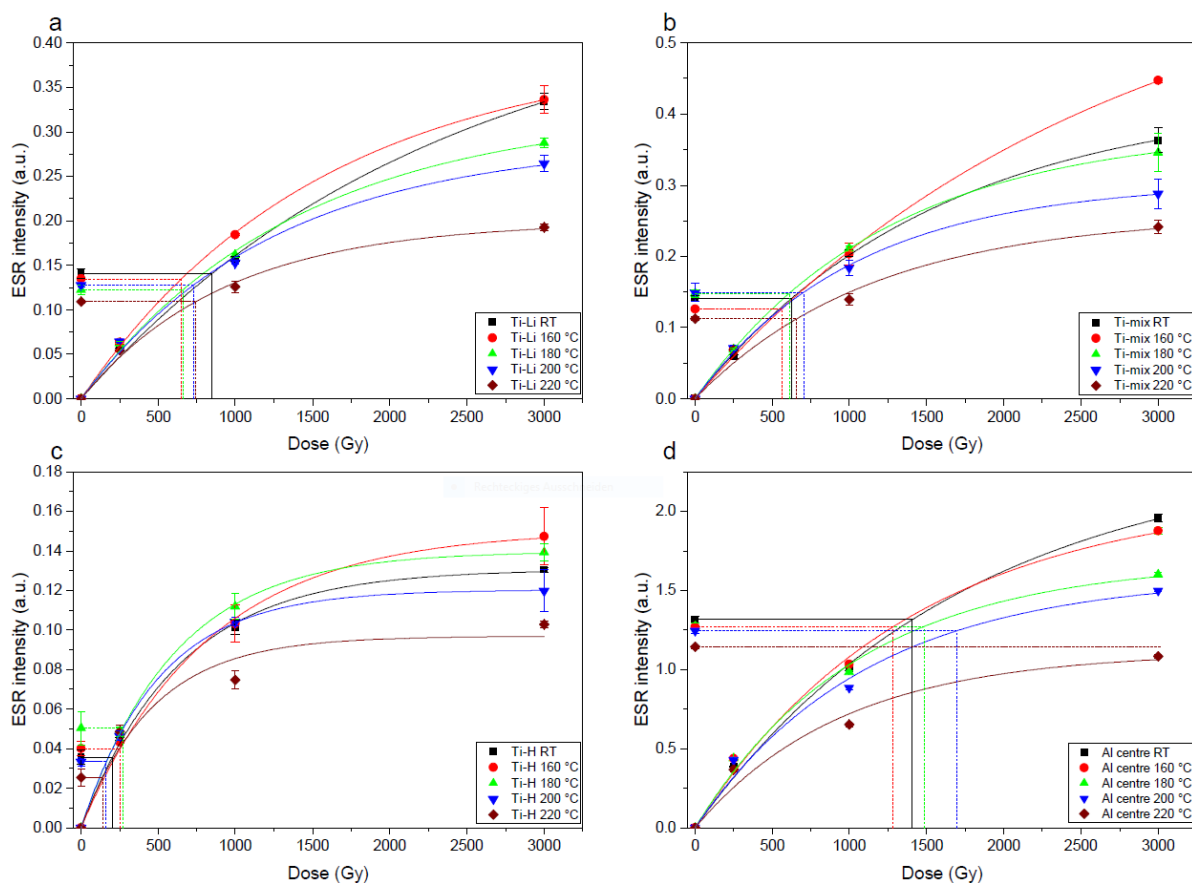
The figure above shows the DRCs for each preheat temperature (RT = room temperature = 20 °C) for each one of the ESR centres. The DRC's for the Al signal (d) show a systematic flattening in the DRC due to lower signal intensities achieved by the same dose applied with a higher preheat

temperature. In case of a preheat temperature of 220 °C the intersection from DRC and the natural signal intensity is > 3000 Gy. This figure has been added to the supplement (Figure A1).

At this point one has to ask himself why one needs to preheat in ESR dating? How come a single type of defect can have a part that is not thermally stable and one that is not, as mentioned in Toyoda and Ikeya, 1994 QG, and explained here as "a preheat plateau test was carried out to assure only stable signals are used"?

We think it is possible that after artificial irradiation (with a few order of magnitude higher than in nature) there can exist thermally unstable Al and Ti centres. Thus it is beneficial at least to conduct a mild preheat to check if it is not the case. Our previous studies (Tsukamoto et al., 2015; 2018) showed the need of preheating.

Especially in our study of the Chinese loess-palaeosol sequence (Tsukamoto et al. 2018, RM) when we compared artificial and natural DRC, that the signal intensity and D_0 of the laboratory ones are much larger than the natural ones. Both decrease with preheat and match better with the natural DRC, confirming that a preheat step is necessary after artificial irradiation.

And how does that relate to the values of E and s quoted in studies such as Richter et al., 2020, QI? ESR is selective, as such only one type of defect is targeted. Does this mean that we acknowledge that both localized and delocalized transitions are at play? If the localized route is the one that leads to an apparent instability, then one should expect the degree of instability to be dose dependent (see Benzid and Timar-Gabor 2020, AIP Advances 075114). If indeed there is a degree of instability that is dose dependent that might explain the observation I am referring to above.

As described in Tsukamoto et al., 2018, we preheated at the same temperature (160 °C, 4 min) for all three aliquots before isothermal measurement started, and the normalised the intensity to the first preheated signal. So we want to make sure that anything thermally very unstable must have gone by the initial preheat.

My last comment concerns the interpretation "The opposite tendency of the heat treatment for the Ti-Li and Ti-H centres in the preheat plateau test (Fig. 2) suggests that some charge transfer between the Ti-Li and Ti-H centres is likely." Personally, I am unable to grasp what model is proposed. Please be more specific. What is the transferred charge?

We do not have enough evidence for a detailed discussion, so we have decided to remove the section cited by the reviewer from the MS.

Technical comment: line 70-reference to table 3 should be made instead of table2.

This reference has been changed to Table 4 since we added another Table in the beginning of the MS.