➔ We thank the reviewer for their positive assessment of our work. We have responded to the raised issues below, where the original comments are in grey font and the responses in black prefaced by an arrow.

Summary

Over the past two decades, two groundbreaking luminescence new dating techniques have emerged: infrared radiofluorescence (IR-RF) on feldspar specimens and infrared photoluminescence (IR-PL). Although IR-RF is the older method, it has only recently gained more adoption. Since the introduction of IR-PL, there has been ongoing debate regarding its superiority over IR-RF. The primary advantage of both techniques lies in their ability to measure the charge density of a principal trap. This measurement provides a more direct estimate of the accumulated dose.

The manuscript presents in essence a long-overdue systematic comparison of IR-RF and IR-PL that should have been done long ago on a set of samples. The present contribution goes even beyond as it also re-analysed older results from a study that has questioned the reliability of IR-RF.

The approach is systematic, and the manuscript is carefully prepared. It takes a somewhat neutral standpoint, although most of the tests are clearly related to IR-RF. Nevertheless, what is refreshing is that it does not claim superiority of one method over the other but tries to point out differences and challenges where applicable.

The manuscript clearly aligns with the scope of Geochronology and should be published.

I have only a few general remarks and more minor technical comments, but I am confident that they can be addressed by the study authors easily.

General remarks

- The manuscript is generally well-structured and well-written, with most sections being easy to read. However, I had the impression that the authors added more experiments along the way and somewhat forgot the original purpose of the manuscript. While this is common, I suggest that the authors revise the introduction to make it easier for readers by clearly marking hypotheses that can be quantified and tested. For instance, the title suggests a straightforward comparison of IR-RF and IRPL, but the study then presents a diverse range of tests (including pIRIR measurements). In other words, the study lacks some rigour and could benefit from a little streamlining.

➔ We have included the IRSL/pIRIR aspect in the introduction, though we highlight that these results are taken directly from the IRPL protocol of Kumar et al. (2021) and not measured separately. Though our focus was on the IRPL results, we chose to also report the IRSL ones, since we had the data.

➔ We have also rephrased the last paragraph of the introduction to make the hypotheses more explicit.

- The MAR test is intriguing, but it comes unexpectedly and lacks further substantiation. This presents a missed opportunity. The authors should either extend this section or remove it, as it will not significantly alter the outcome but blur the story. By doing so, they would also have the opportunity to design more robust experiments and incorporate modelling.

➜ We agree that the IR-RF MAR results are preliminary and that more work still needs to be done. However, we believe that testing a MAR approach follows logically from the reported issue of uncorrected sensitivity changes in the SAR. We have chosen to retain the section to serve as a motivation for future research. We state in the conclusion that the MAR results are preliminary and have now expanded on the reasons behind the inclusion of a MAR approach in the introduction, so it does not appear as unexpectedly.

- The same goes for the discussion about the detection window. This is given quite some attention, but at the end, it seems to have no effect. If so, this can be shortened to two or three lines and with the rest of the data in the supplement.

➜ Agreed, we have removed the detailed descriptions from the main text and kept the results in Fig. S6 and Table 4.

- The authors provide access to raw data, which is highly appreciated; however, they should also add full information on the source rate calibration (see below). Otherwise, the results are of limited use to others.

➜ We have added the dose rate calibration details used for the new measurements. For the data from Buylaert et al. (2012), the dose rate used for each aliquot is provided in the Zenodo dataset under 'RLanalyse_De.csv'.

- The definition of uncertainties appears somewhat ambiguous, often I am not sure whether they are truly comparable. For instances, sometimes it seems to be the standard deviation, sometimes the standard error of the mean, for the independent age control something else(?).

➜ The stated uncertainties for our mean D_e estimates are the standard error of the mean. This metric was chosen to remain comparable with the work of Buylaert et al. (2012) and the independent ages reported therein, whose uncertainties according to their Table S1 represent one standard error. In the few cases in which the standard error of the mean was below the channel length (10 s, approximately 0.6 Gy), we considered the channel length as the mean D_e uncertainty because that is the limit of our resolution. We have added this information to section 2.3. To avoid confusion, we have removed the D_e estimate of one aliquot from the caption of Fig. 1, whose uncertainty (the standard deviation of a Monte Carlo simulation of the D_e distribution bootstrapping the residuals of the curves after sliding) was not used for later calculations.

Minor comments

- L37: Because you explain basics, you should also addd a suitable reference for IRSL

➔ Agreed, we have added a reference.

- L38: The reference to Krbetschek et al. (2000) seems incorrect. The authors wrote "Fading tests (storage over periods of several months at room temperature) have shown signal stability." Krbetschek et al. (2000, p. 497). They further stated: "Further investigation is necessary to ascertain what this tells us about the mean life of the trap population" (2000, p. 497). They did not write anything about "lower anomalous fading rates". They wrote about signal stability and the mean life(time) of the trap population. This implies that they meant the thermal not the athermal lifetime.

➔ We interpreted the result of "signal stability" (Krbetschek et al., 2000; p. 497) obtained from the fading test as a fading rate consistent with zero, which is lower than the usual non-zero rates obtained for IRSL. In any case, we have rephrased this sentence to "a more athermally stable signal".

- L46: Technically, the sliding approach goes back to Prescott et al. (1993) (or even earlier) under the name "Australian slide". Buylaert et al. (2012) describe horizontal sliding in their Fig. 4; the method/tool is described in Lapp et al. (2012) where they describe a time-shift. Kreutzer et al. (2017) (see also in Murari et al. 2018) first used the approach; Murari et al. (2018) formally introduced it. However, Buylaert et al. (2012) indeed mention horizontal and vertical adjustments, but it is unclear what their conclusion was and why they did not test it. But I agree, credit should be given to them because they mention the idea.

➔ Agreed, we have rephrased and added the reference Kreutzer et al. (2017).

- L86-L95: This paragraph is very muddled. You start re-analysing 16 samples but present 10 new IR-RF and then again "eight samples originally used". Please rephrase to improve readability or make a list for your experiments or a workflow graph.

➔ We have rephrased the paragraph and added a supplementary overview figure showing how our chosen set of samples relates to those measured in previous studies for clarification.

- L86-L95: The introduction should explicitly state a research hypothesis that will be tested in the contribution, rather than presenting a list of experiments that may or may not yield a specific outcome; some of them even unrelated to the study title.

➔ The hypothesis behind each of the tests is that varying that parameter will lead to more accurate IR-RF D_e values. We have rephrased the final paragraph of the introduction to clarify this and name the tested parameters, e.g., the use of the vertical slide, the length of the natural dose curve used for sliding, the number of rejected initial channels, the detection window and the type of bleach between the natural and regenerative dose curve measurements.

- L102-L103: The HF etching of feldspars is challenging, and still it is unlikely to remove any other alpha-irradiated layer uniformly and in the desired manner (Duller, 1992; Porat et al., 2015; Duval et al., 2018). Given that sample preparation cannot be altered retrospectively, I propose discussing this issue and its potential impact on the final results later in the manuscript.

➔ We have commented on this issue in section 2.1, but since all samples expected to be in the dating range received the same treatment, we do not expect etching to affect our conclusions.

- L116-L125: Please provide detailed information on the calibration of the other machines, including the aliquot size, sample carrier, dose rate, and calibration date. This information is essential for cross-checking your results by others without the need for additional inquiries. If necessary, please indicate whether you had to correct calibrations based on previous measurements (as described in Autzen et al., 2022). I attempted to recalculate a few results using the data provided on Zenodo. However, without the dose rate (available for some samples, though), I am unable to effectively compare the results.

➔ We have added the dose rate calibration details for the missing readers used for the new measurements.
➔ We have now corrected all measurements taken from Buylaert el al. (2012) using the 8.25% value suggested by Autzen et al. (2022). We also used this value to correct the control ages that had been estimated using OSL.

- L136: Please state the number of ignored channels, or the dose (you do that later). I tried to recalculate, for instance, A8. But my result is consistently 0, however, I can get any result (also the one you report) by ignoring a certain number of channels. Means, this information matters.

➔ We did not state the number of rejected channels here in the methods section because this was a parameter we varied. We have added a sentence stating that between 0 and 499 channels were ignored, which will be detailed in a later section. In Table 4, where we present results from the optimal tested parameter combination, we detail in the caption that 2 Gy were removed. We have added the number of ignored channels to the caption of Table 4 as well as to the discussion.

- L139: A more detailed description with of the initial rise can be found in Frouin et al. (2017) (their supplement). From this analysis, it becomes evident that the response appears to be dose-dependent and exhibits a range of responses within a given dose range.

➔ We now refer to the initial rise results from Frouin et al. (2017).

- L163: What is the justification for the double-exponential fit? Wouldn't the GOK model (Guralnik et al., 20215) be a better candidate for feldspar?

➔ We used a double exponential to be able to directly compare our results with those from Kumar et al. (2021). We do not expect that a GOK model would change the results significantly.

- L176: Can you show such a distribution?

➔ We have added example D_e distributions (850/40 nm filter) as a supplementary figure.

- L215: I think that the number of channels matters more than the dose; please also check the supplement by Frouin et al. (2017) where this investigated (although it seems only with horizontal sliding).

➔ We have checked the suggested investigation by Frouin et al. (2017), who undertook an analysis of incrementally increasing the segment length up to 100 channels (horizontal slide). Our results are compatible with those presented there when we apply only the horizontal slide, as shown below for sample H22553 (always rejecting the first 4 channels). Looking only at the initial 100 channels (Fig. inset), it appears that a plateau is reached after 60 channels (~36 Gy). However, if continuing the analysis, the mean D_e rises again and continues to change until ~2500 channels (~1500 Gy), then reaching a plateau, which extends until at least 3000 channels (~1800 Gy). The early plateau is reached here at a similar dose as in the study of Frouin et al. (2017), who report >40 channels (~30 Gy with their 0.7 Gy/channel) are needed to reach a D_e plateau. Overall, it seems that a segment length of <100 channels does not yet tell the whole story. We now refer to the results from Frouin et al. (2017) in a new section "Considerations on the IR-RF DRC".
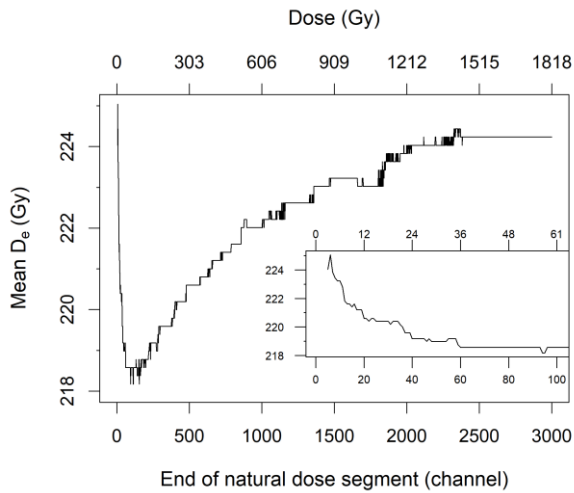
➔ To answer whether the channel length or dose is the more relevant measure, we have re-measured one aliquot of sample 072255 changing the channel length from 10 s to 1 s. These settings serve as a proxy to running the sequence on a reader with a dose rate 10x lower. As shown below, the mean D_e follows the same pattern when looking at the dose (panel a), but not so when looking only at the channels (panel b). This suggests that the cause of the pattern is a dose-dependent characteristic. We have added the number of channels in some key places in the manuscript for reproducibility, but highlight that users would need to adjust measurement parameters (e.g., curve length) to obtain equivalent results when using devices with different dose rates.
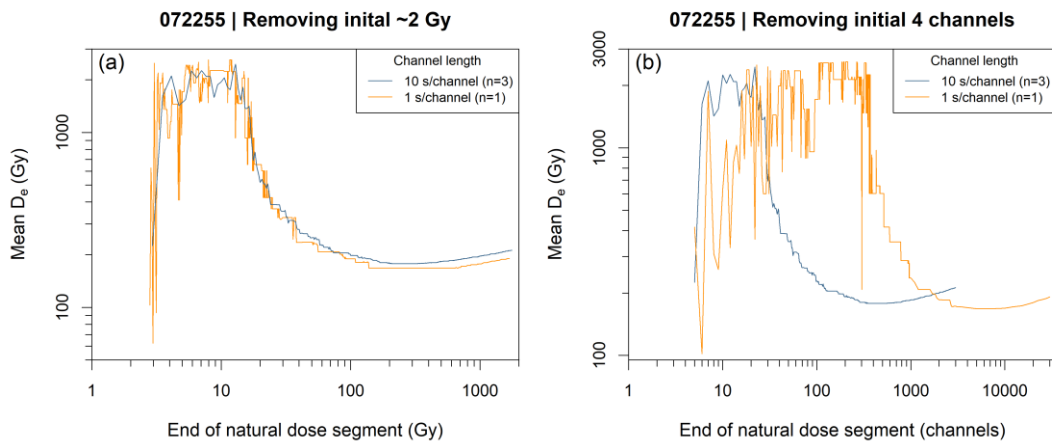
- L430: I concur with this conjecture, albeit with a slight distinction. I think that the relationship lies not solely with the dose but also with the number of channels. Your objective should be to identify a plateau of equivalent dose values rather than distinct segments. To achieve this, you can segment your natural dose and incrementally add channels to the RF natural dose until a plateau is detected. While this approach addresses the issue of channel-related variations, it still presents a challenge: if the regenerated and natural curves indeed differ, the results may be inaccurate when compared to an independent age control.

  Nevertheless, this approach eliminates the possibility of arbitrary channel selection.

➔ See comment above on the issue of channels vs. dose.

➔ We have tested the reviewer's suggestion of running the segment length comparison incrementally adding one channel as opposed to our original approach of selecting six segment lengths. Results from one sample are shown below for (a) horizontal and (b) horizontal and vertical slide. Despite testing with a relatively long natural dose curve of 1800 Gy (3000 channels), we do not yet see a D_e plateau with the horizontal and vertical slide. In fact, there is an initial decrease (possibly related to the sliding algorithm) and then an increase. For this sample, the D_e is beginning to stabilize around 1800 Gy, but for other samples the increase is still significant at the same segment length. We have chosen six segment lengths (including a new one not investigated in the original manuscript) that cover the range of likely D_e values (magnification of the region of interest given in the inset). The chosen lengths are also shown in the figure below (coloured points).
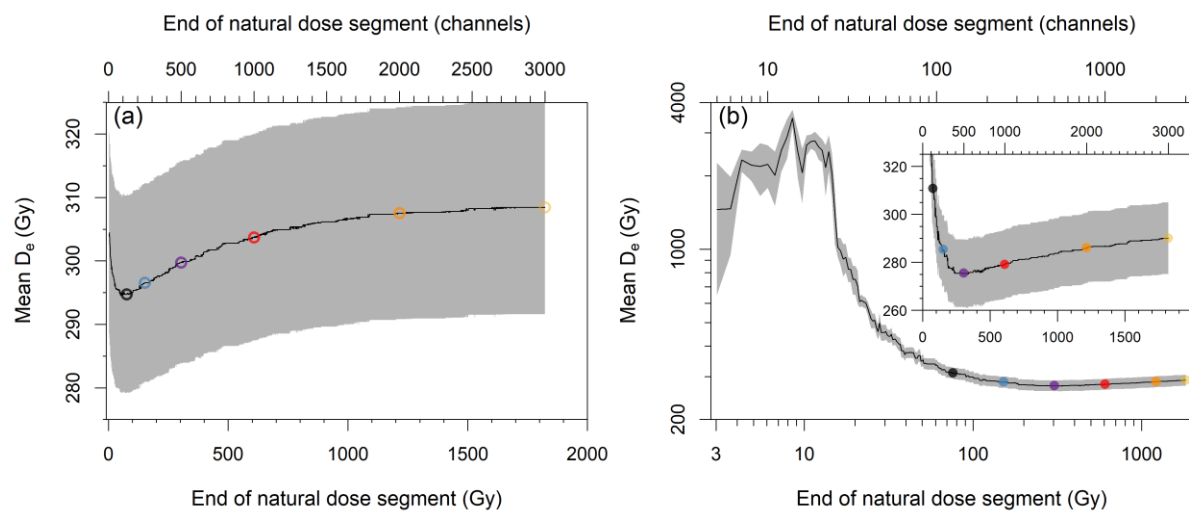
- L440-L443: They yield 0 Gy because the algorithm has no other choice to match the curves given the shapes and the starting points and then sets it to 0. This is not coincidental; this is by design. See your own arguments a few lines below.

➔ We meant that it is a coincidence that when not rejecting any initial channels the expected dose is obtained. The agreement should be considered to be an analytical artefact (due to the initial rise always being at the start of the curves) and not evidence of accuracy. For this reason, we recommend always rejecting initial channels even though that leads to a worse accuracy in the case of the modern samples. We have re-phrased the sentence for clarity.

- L437: This somewhat contradicts your chain of arguments trying to emphasise good arguments and put more weight on one or the other. The 35 Gy is an arbitrary choice and sample dependent; it seems a good fit for your samples, but I suggest refraining from generalising this observation. The best approach seems to reject the very first channel and keep the rest (with a certain number of minimum channels)

➔ We did not recommend rejecting 35 Gy, instead, since for some samples the D_e is not stable when rejecting more than ~35 Gy, no more than that should be rejected. In fact, for most samples, rejecting much less is sufficient, as evidenced at the end of this sentence, where we state we rejected only ~2 Gy. We have rephrased for clarity.

- L461-L464: In Murari et al. (2021) all measurements (Risoe and Freiberg readers) seem to have been used 70ºC as recording temperature; please rephrase or remove.

➔ We were referring to the room temperature RF measurements of Buylaert et al. (2012), but we have removed the indication of temperature in the sentence for simplicity.

- L465-L470: Agreed, but you should also point to the different protocols with no less than 17 to 18 steps. I am wondering how sensitive the equivalent dose is to certain parameters. If you cannot test this, you should at least discuss it.

➔ Agreed, added to discussion.

- L483-L484: I do not believe that the comparison to the quartz model is valid. While the observation may share some similarities, the underlying mechanism is unknown and likely distinct in quartz. Unless you can provide a model and demonstrate that the mechanism is indeed similar, I recommend removing this speculative comparison. The subsequent comparison is more appropriate and logical, although it is purely descriptive.

➔ Agreed and removed, though we added a sentence on previous K-feldspar MAR studies.

Figures

- Figure : The figure I am missing is a distribution plot for equivalent doses. Perhaps this can be added for suitable samples.

➔ Added as supplementary figure.

- Figure 1: Please add information on the aliquot number and which measurement window was used. Also, here contrary to what was written in the M&M section no initial channel was discarded.

➔ We have added the requested information and a note on the channels used for sliding to the caption. In this case, the initial 4 channels were discarded for sliding. Since the

aim of this figure was to introduce the sliding procedure to readers unfamiliar with it, for simplification, we did not use different colours for the discarded channels. The concept of rejecting certain channels is later introduced graphically in Fig. S4.

- Figure 2: Please colour-code the samples and use shapes to denote the methods. This will prevent readers from having to guess which sample is shown. If you run out of distinguishable colours, please use labels. For the 600 Gy exposure, the quantity of channels is the primary parameter of concern, rather than the dose (the information remains beneficial regardless).

➔ We have changed Fig. 2 using colours to distinguish the samples but have retained a method-based colour scheme in Fig. 5, since the number of samples is lower there, allowing them to be distinguished by their dose (or by comparison with the new figure version).
➔ Given the relatively long segments we are working with, we don't expect the observed differences in D_e to be caused by the sliding algorithm (in which case the number of channels might be the primary parameter), but to represent true sample characteristics related to the dose-response, i.e., irrespective of channel length or reader dose rate. See also our reply to the comment on line 215. However, for completion, we have added the number of channels to the caption.

- Figure 3: What is the central new information conveyed by these figures? The sliding method, particularly requires offsetting for short segments and less curvature. Please condense to a single key figure with a succinct message.

➔ This figure is intended to be descriptive and helpful for readers who prefer a visual representation rather than, e.g., tabular data presentation. The main messages described in section 3.3 are (i) depending on the segment used for sliding, there can be variation in the resulting mean D_e and (ii) this behaviour is sample-dependent. The wider implications are then summarized in Fig. 9: for samples with small changes in D_e, the conventional SAR approach can be expected to yield accurate D_e values. We have added a second panel to Fig. 9 summarizing the change in D_e for increasing segment lengths.

- Figure 5: To compare IR-RF and IRPL, it is necessary to include a third figure that compares both techniques with the method you believe performs most effectively. Additionally, you should compare the distribution of the relative residuals from the 1:1 regression line to assess whether there is a significant difference between the two methods or if they are merely random.

➔ We have added a figure using the APh-IRPL$_{880}$ results in the x-axis.
➔ We have tested the relative residual comparison suggested by the reviewer (for the non-modern samples that have expected ages). The distributions are relatively similar for the three methods, however, we believe that the sample size is too small to reach a definitive conclusion especially considering possible dose-dependent differences. For this reason, we would not include it in the manuscript but only in the supplement.

- Figure 6: Do the grey bands make sense? The optimal range would be with the highest number of channels. However, if you wish to retain the current settings, you should also experiment with different integration values for the other methods. But I guess then it becomes very confusing.

➔ We agree with the reviewer that the equivalent comparison would also include the integration values for IRPL/IRSL, but also think that this inclusion would make the already complex figure somewhat convoluted. We included the grey bands to highlight the subjectivity of our chosen number of channels, which we show in section 3.3 to have a significant effect for some samples. As shown in Fig. 3f, increasing the number of channels (until reaching ~1800 Gy) does not lead to a D_e plateau and we caution against the assumption that a higher number of channels will necessarily be optimal for dating.

- Figure 9: It requires an illustration of the separated dose signal components. Currently, it appears a little bit arbitrary and descriptive.

➔ We have added a legend showing the different signal segments.

- Figure S4: The offsets of the curves are a little bit difficult to see, perhaps you can show the residuals?

➔ The residuals were added to Fig. S4 and S5.

- Figure S9: What does this figure add to the manuscript? Your concern is the comparison of two methods, here you compare all kind of protocols and procedures on top of two types of IR-RF and IRPL. I can somewhat understand your Fig. 6 in the main text, but this seems too much.

➔ This figure only compares how results from two protocols plot against the expected values: in panel (a) the IRPL protocol (which contains sequential IRSL and IRPL steps) and (b) the IR-RF protocol. We have added a sentence to the caption to clarify this. Our IRPL measurements were undertaken to increase the dataset presented by Kumar et al. (2021), so we used their MET pIRIR-IRPL protocol. Since our study's focus is the comparison of IR-RF and IRPL, the IRSL D_e are not discussed in the main text, but we show them in the supplement so as not to ignore the data.

Tables

- Table 1: Instead of 'se' that refers to the standard error (of what?), please use confidence intervals.

➔ The expected ages and their associated uncertainties (standard error of the mean) are taken directly from the papers in which they were dated, so we have kept them to remain comparable with the previous work.

- Table 3: What do the uncertainties represent? The standard deviation? For such a low 'n' you should rather calculate confidence intervals using the *t*-distribution unless you can show that the normal distribution approximation is valid.

➔ The uncertainties of the D_e estimated in this work are the standard error of the mean, which has been added to the caption. Whereas we agree that the t-distribution confidence intervals would be statistically more accurate, we chose to retain the standard error to stay comparable with the studies against which we compare our results, i.e., Buylaert et al. (2009) and Kumar et al (2021). We have calculated the confidence intervals for the main data set and confirm that the same samples match the expected ages, so our conclusions are not affected by the choice of distribution. We have also added the confidence intervals to the new KDE plots in the Supplement. Since the number of aliquots are given for each estimated D_e, interested readers have all the parameters to calculate confidence intervals for the other data sets.

- Table 4: Please explain the meaning of the uncertainties and align them. I suggest calculating consistently 95% confidence intervals.

➔ We have added to the caption that the uncertainties represent the standard error of the mean. As for the suggested confidence intervals, see our response above.

References

Autzen, M., Andersen, C.E., Bailey, M., Murray, A.S., 2022. Calibration quartz: An update on dose calculations for luminescence dating. Radiation Measurements 106828. https://doi.org/10.1016/j.radmeas.2022.106828

Buylaert, J.P., Jain, M., Murray, A.S., Thomsen, K.J., Lapp, T., 2012. IR-RF dating of sand-sized K-feldspar extracts: A test of accuracy. Radiation Measurements 47, 759–765. https://doi.org/10.1016/j.radmeas.2012.06.021

Duller, G.A.T., 1992. Luminescence chronology of raised marine terraces, south-west north island, New Zealand (PhD thesis). University of Wales, Aberystwyth.

Duval, M., Guilarte, V., a, I.C. n, Arnold, L.J., Miguens, L., Iglesias, J., lez-Sierra, S.G. a, 2018. Quantifying hydrofluoric acid etching of quartz and feldspar coarse grains based on

weight loss estimates: implication for ESR and luminescence dating studies. Ancient TL 36, 1–14. https://doi.org/10.26034/la.atl.2018.522

Erfurt, G., Krbetschek, M.R., 2003. IRSAR - A single-aliquot regenerative-dose dating protocol applied to the infrared radiofluorescence (IR-RF) of coarse-grain K-feldspar. Ancient TL 21, 35–42. https://doi.org/10.26034/la.atl.2003.358

Frouin, M., Huot, S., Kreutzer, S., Lahaye, C., Lamothe, M., Philippe, A., Mercier, N., 2017. An improved radiofluorescence single-aliquot regenerative dose protocol for K-feldspars. Quaternary Geochronology 38, 13–24. https://doi.org/10.1016/j.quageo.2016.11.004

Guralnik, B., Jain, M., Herman, F., Ankjærgaard, C., Murray, A.S., Valla, P.G., Preusser, F., King, G.E., Chen, R., Lowick, S.E., Kook, M., Rhodes, E.J., 2015. OSL-thermochronometry of feldspar from the KTB borehole, Germany. Earth and Planetary Science Letters 423, 232–243. https://doi.org/10.1016/j.epsl.2015.04.032

Krbetschek, M.R., Trautmann, T., Dietrich, A., Stolz, W., 2000. Radioluminescence dating of sediments: methodological aspects. Radiation Measurements 32, 493–498. https://doi.org/10.1016/s1350-4487(00)00122-0

Kreutzer, S., Murari, M.K., Frouin, M., Fuchs, M., Mercier, N., 2017. Always remain suspicious: a case study on tracking down a technical artefact while measuring IR-RF. Ancient TL 35, 20–30. https://doi.org/10.26034/la.atl.2017.510

Lapp, T., Jain, M., Thomsen, K.J., Murray, A.S., Buylaert, J.P., 2012. New luminescence measurement facilities in retrospective dosimetry. Radiation Measurements 47, 803–808. https://doi.org/10.1016/j.radmeas.2012.02.006

Murari, M.K., Kreutzer, S., Fuchs, M., 2018. Further investigations on IR-RF: Dose recovery and correction. Radiation Measurements 120, 110–119. https://doi.org/10.1016/j.radmeas.2018.04.017

Murari, M.K., Kreutzer, S., Frouin, M., Friedrich, J., Lauer, T., Klasen, N., Schmidt, C., Tsukamoto, S., Richter, D., Mercier, N., Fuchs, M., 2021. Infrared Radiofluorescence (IR-RF) of K-Feldspar: An Interlaboratory Comparison. Geochronometria 48, 95–110. https://doi.org/10.2478/geochr-2021-0007

Peng, J., Wang, X., Adamiec, G., 2022. The build-up of the laboratory-generated dose-response curve and underestimation of equivalent dose for quartz OSL in the high dose region: A critical modelling study. Quaternary Geochronology 67, 101231. https://doi.org/10.1016/j.quageo.2021.101231

Prescott, J.R., Huntley, D.J., Hutton, J.T., 1993. Estimation of equivalent dose in thermoluminescence dating - the Australian slide method. Ancient TL 11, 1–5. https://doi.org/10.26034/la.atl.1993.204

Porat, N., Faerstein, G., Medialdea, A., Murray, A.S., 2015. Re-examination of common extraction and purification methods of quartz and feldspar for luminescence dating. Ancient TL 33, 22–30. https://doi.org/10.26034/la.atl.2015.487