Point-by-point reply to reviewers

Reviewer 1

the use of several gamma doses for beta source calibration and then taking the regression of a plot of gamma dose vs. recovered beta dose to derive the dose rate of the beta source. This is an approach that should certainly be promoted, and hence deserves a bit more weight in the manuscript now added in the abstract

l. 22: What do you mean with "geometrical function"? Maybe just replace by "irradiation geometry"? Done

l. 39: What is meant with the "interplay between sample and sample carrier"? Please be a bit more specific here. Replaced by *"...*on the atomic numbers (*Z*) of mineral and sample carrier (line 40)

l. 75 (Table 1 caption): it should read ":: are derived from MC simulation". Done

While readers can look up in Hansen et al. (2015) for the DTU quartz, nothing is written here about the Freiberg quartz. The Freiberg quartz is now published, and the reference is provided in line 83.

Please increase the font size in Fig 1 – done.

Which stimulation power density was used for which aliquot size and why? Please provide the reasoning why you chose this approach? The revised text says that the stimulation power was changed to avoid overexposure of the photomultiplier …depending on the size of the aliquot (header of table 2)

Table 2, last row: it should be "R108_4" – yes, done

Does "depth of dose rate" mean "dose rate as a function of depth"? Consider re-phrasing. Re-worded following the suggestion (line 120)

The units given for the simulated layers of the sample "cylinder" should be _m instead of mm, I would think (same in the caption to Fig. 2). Yes, and sorry for this formatting error.

l. 130: Does this statement in brackets mean that the dose is registered in Gy for each starting particle, i.e. particle emitted from the source? In the source code the results of the F6 tally are reported in dose per source particle (Gy) averaged over the target cell. Because MCNP results are normalised in our study the unit of the F6 tally is not relevant, we have removed the unit in the text (line 141). In the supplement 'number of tracks' are reported – these are a direct output of the code.

write one or two additional sentences on the specific purposes of the GEANT4 and MCNP6 codes, i.e. which code was used for which part of the simulation. Done (lines120 and 140)

Improve Table 3: sample code, number of digits. Done

*Line 157 - How does the "total uncertainty of experimental data" of 5-8% relate…*1 and 2 sigma errors were used for comparing data, but because *n* (number of aliquots measured) is low, these are actually not statistical errors suitable for comparison. We have re-calculated all errors and listed

these in Table 3. Obviously, errors are big and would only become small when increasing *n*. On the other hand, the accuracy of the value will not change with *n*>20 due to the excellent reproducibility of the calibration quartz. Thus, error bars are not plotted in Figs 3 and 4.

It should read ": : : quoted at… Done

Table 4: I suggest expressing the difference in dose rate between different grain size fractions not in percent, but as a ratio - done

consistency of axis labelling – thanks yes, done.

Why are the GEANT4 simulations skipped for grain sizes >250 _m, while they were carried out for the MCNP6 simulation? - This was simply motivated by workload of the respective expert and/or the demand on the computational resource.

Fig. 4: This figure seems to be identical with Fig. 3a. Please check and update. Done

Black and red dots in the caption (legend to plot) should be swapped. Done

Fig. 7 - What is the purpose and meaning of the simulation without sample holder? Can inferences be made about the role/magnitude of electron backscatter from the sampleholder? Maybe this aspect should be shortly discussed in the manuscript. Explanation now added (lines 230)

Fig. 8: If the dose rate is shown normalized to the 10 _m large aliquot simulations, why do the data start at _108% in the center of the sample carrier? The figure shows the dose rate profile for various grain- and aliquot sizes. It does not show the average dose rate. The normalised average value of the 10 µm curve is at 100%. Now explained in the figure caption.

What is "purpose-prepared sample material"? – it is a natural sample (e.g. dune sand) prepared for the purpose of becoming beta-source calibration material. Here it is annealing and repeated irradiation and read-out using blue-light stimulation in order to sensitise the quartz.

Reviewer 2

We acknowledge the need for saying more about source uniformity in the text. A description of the sources in terms of (unknown) homogeneity and shape of the radiation field is now added (lines 69ff).

It would be very helpful to the reader if this information was given for each instrument and source used in the present study, since it would allow more meaningful interrogation of the data presented. For example, does the "open ring" have a larger active face relative to the size of the sample holder than for other sources? If so, this might explain why aliquot size does not appear to be an important consideration for this source. Done (lines 69ff).

The number of aliquots used to produce each datapoint in the experimental datasets is often rather low. For example, only two 8 mm diameter aliquots of R108_4 were measured using the lexsyg

SMART closed ring source….. I accept that the calibration samples used are bright and highly reproducible, but reliance on very small numbers of aliquots makes the resulting dataset prone to other sources of uncertainty e.g. non-uniform aliquot preparation, grain(s) between the sample and hotplate, deformed sample holders (Duller et al., 2000) etc etc. The authors should at least comment on why they think that the small number of aliquots measured doesn't pose a problem.
Uncertainties listed were actually not statistical errors suitable for comparison. We have re-calculated all errors and listed these in Table 3. Obviously, errors are big and would only become small when increasing *n*. On the other hand, the accuracy of the value will not change with *n*>20 due to the excellent reproducibility of the calibration quartz. Thus, error bars are not plotted in Figs 3 and 4.

Three aliquot diameters were measured for sample R113_180 using the lexsyg SMART closed ring source, yet only data for the 5 mm diameter aliquots are provided in Table 2. Apologies – correct numbers now added.

In Figure 3a, the caption states that all data are for the 8 mm diameter aliquots, yet this size wasn't measured for sample F14_90 using the lexsyg SMART closed ring source, and the data presented appear to be for the 5 mm aliquots. Thanks – now clarified in the caption

Similarly, the data in Figure 3b are normalised to the 8 mm diameter aliquots, but I suspect that the 5 mm datapoint for F14_90 closed ring source is the 5 mm data normalised to itself, which is misleading. The caption now states that data were normalised to 8mm and 5mm aliquot data.

Figure 4 is the same as Figure 3A – apologies. The correct figure is now inserted.

The data in Figure 7 claim to be normalised to the 10m grain size and 7.95 mm (reported as 8 mm elsewhere in the paper) aliquot diameter, whereas I suspect they are normalised to the 10 m, 5 mm aliquot
We use 7.95 mm aliquot size for experimental and simulation data because this is the inner diameter of the cup (minus rim) which is entirely covered for fine grained samples. '8mm' is now removed from the text

why (and how?) did you perform the GEANT4 simulation without the sample cup?
We have changed the text (lines 230f)

don't really understand the paragraph starting on line 234, which is problematic because it appears to be critical to your explanation of grain size effects with small aliquots. Please expand and clarify.
We have changed the text (line 250ff).

Line 65: The phrase "ring shaped source closed to the top" is awkward.
We have amended the text in sec 2.1 where sources are described. The diagrams are available through the references.

Figure 1. Please make the small text larger. Done

Line 99: Please explain the rationale for using different stimulation powers for different sized aliquots and show which power was used for which size. Now explained in the table header

If the reduced density [1.8 g cm^{-3}] is a modeller's approximation of sand deposited as a monolayer on a sample holder, please explain the logic yes, this is correct. We have changed the text accordingly (lines 130-122)

Figure 2: Please add A and B to the figures. Done

Please also provide a clearer explanation of exactly what Figure 2B is simulating – Fig. 2B shows seven grains in plan view; they represent spheres of SiO_2 as stated in the caption

Line 153: …but it would be better to be specific e.g. 4-11 µm, a text is now added which describes the terms 'coarse' and 'fine' grain aliquot (lines 79ff)

move the "(fg=fine grain)" statement somewhere else since it is currently out of place. Thanks, done

Table 3 and 4: Dashed and solid horizontal lines appear more or less at random. Please standardise Done.

you state that for 8 mm aliquots the effect of grain size is insignificant. By what criterion? If by your own stated in Line 171, please say so. Not sure what is meant here. The previous paragraph does outline the criteria for significance

Line 181: "6-26%" should read "0.6-26%". Thanks, it should actually be 0.4 % (line 192)

Line 182: ". . .the magnitude of the difference is controlled by the shape of the source". This sentence requires more explanation. We feel the sentence is clear: the magnitude is the size of the difference and this size is small for the open-ring source (e.g., 0.4%) and big for the closed ring source (26%)

Sentence starting on Line 182…: now clarified with the correct Fig 4 displayed.

I'm not convinced that the Armitage and Bailey (2006) data "jump" between 50 and 100 µm .. A&B (2005) themselves identify a significant difference and say: "…*with the dose rate to 4–11 µm grains being ~12% lower than that for 55–250 µm grains*." We call this 12% difference a "jump".

As the authors suggest (line 191), the experimental dataset is actually rather similar to the simulated data. This is not quite what we say. Our sentence is : "There is a striking similarity between the simulated data and the experimental data adopted from Armitage and Bailey (2005), but the simulation shows a gradual change of the grain-size effect, while the experiment indicates a "jump"…"

Line 224: I think this should read ">5%" not "<5%". No, <5micron is correct. The drop of dose rate when grains are big (>200 micron) and aliquots are small (<5mm).

Line 239: The figure reference should probably be to Fig. 9, in which case a reduction in the absorbed dose of ~3% looks more appropriate by eye. The Fig reference was actually correct, but the sentence was incomplete. The text has changed (lines 250ff).