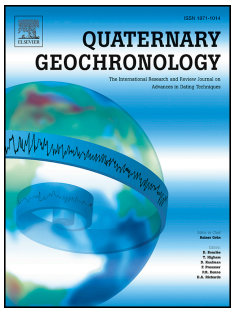
Testing the accuracy of single-grain OSL dating on Eemian quartz samples

Frederik H Baumgarten, Kristina J Thomsen, Guillaume Guérin, Jan-Pieter Buylaert, Andrew S Murray



Testing the accuracy of single-grain OSL dating on Eemian quartz samples

Frederik H. Baumgarten, Kristina J. Thomsen, Guillaume Guérin, Jan-Pieter Buylaert, Andrew S. Murray

PII: S1871-1014(24)00106-7

DOI: https://doi.org/10.1016/j.quageo.2024.101602

Reference: QUAGEO 101602

To appear in: *Quaternary Geochronology*

Received Date: 15 November 2023

Revised Date: 13 July 2024

Accepted Date: 13 July 2024

Please cite this article as: Baumgarten, F.H., Thomsen, K.J., Guérin, G., Buylaert, J.-P., Murray, A.S., Testing the accuracy of single-grain OSL dating on Eemian quartz samples, *Quaternary Geochronology*, https://doi.org/10.1016/j.quageo.2024.101602.

Testing the accuracy of single-grain OSL dating on Eemian quartz samples

Frederik H. Baumgarten[a]*, Kristina J. Thomsen[a], Guillaume Guérin[b], Jan-Pieter Buylaert[a], Andrew S. Murray[a,c]

[a]Department of Physics, Technical University of Denmark, DTU Risø Campus, Roskilde, Denmark

[b]Université de Rennes, CNRS, Géosciences Rennes, UMR 6118, 35000 Rennes, France

[c]Nordic Laboratory for Luminescence Dating, Department of Geoscience, Aarhus University, Aarhus, Denmark

Abstract: Single-grain OSL dating of quartz is a popular approach to OSL dating, even when incomplete bleaching is not likely to be significant. However, little testing of the accuracy of single-grain dating has been published; particularly for samples older than 50 ka. In this study, we investigate the accuracy of single-grain quartz OSL dating, when a significant number of individual grains are no longer able to accurately measure the burial dose because of saturation effects. We compare standard multi-grain OSL results with those obtained from single-grain OSL measurements for five OIS substage 5e (Eemian) samples (~128 ka). We show that for these samples, standard multi-grain quartz dose estimation results in dose estimates in good agreement with the predicted doses (four of the five samples recover age control), but that standard frequentist single-grain dating procedures significantly underestimate the age controls, i.e. the measured to predicted dose range between 0.42±0.03 and 0.84±0.06, where the underestimation increases with increasing relative number of grains in saturation. Attempting to remove the inevitable bias in the dose estimation resulting from a significant number of saturated grains (by using the D_c criterion) reduced the underestimation, i.e. the measured to predicted dose range between 0.63±0.05 and 0.94±0.08, but only the sample with the smallest absorbed dose is consistent with the age control. Using Bayesian analysis ("BayLum") the ratio of measured to predicted dose range between 0.75±0.07 and 1.14±0.08, but only two of the five samples agree with the independent age control. Our results have implications for the evaluation of single-grain OSL dating of quartz in the 100-200 Gy natural dose range.

Keywords: Luminescence, quartz, OSL, single-grain, accuracy, multi-grain, independent age-control, dose models

1 Introduction

Quartz single-grain (SG) OSL dating is often considered to be a superior alternative to multi-grain (MG) dating and is routinely used for dating OSL samples (e.g., Jacobs et al., 2008; Armitage et al., 2011; Arnold et al., 2013; Jacobs and Roberts, 2015; Demuro et al., 2019; Liu et al., 2022; Demuro et al., 2023). Single-grain measurements allow an evaluation of each individual grain and enable the removal of those which are deemed undesirable - either because they show aberrant OSL characteristics (e.g., Jacobs et al., 2013; Demuro et al., 2023), or because they can be identified as sources of contamination from e.g., post-depositional mixing (e.g., Roberts et al., 1999; Fu et al., 2019). In standard multi-grain OSL dating, all grains emitting OSL contribute to the OSL signal. If a sample suffers from a significant degree of incomplete bleaching, the signal averaging in the multi-grain approach will lead to overestimation of the true burial age and hence it is argued that the single-grain approach should be applied to obtain an accurate burial estimate (e.g., Duller, 1994; Olley et al., 1998) through statistical modelling to determine the grain population most likely to have been well-bleached at burial, e.g., the Minimum Age Model (MAM, Galbraith et al., 1999), the Internal/External Uncertainty consistency criterion (IEU, Thomsen et al., 2007) or a Bayesian mixture model (Christophe et al., 2018). But do single-grain measurements also give the most accurate burial dose when incomplete bleaching is not likely to be significant? It seems intuitive that analysis of a subset of grains formed exclusively by grains displaying “good” OSL characteristics should, as a minimum, perform on par with the full set (as used in multi-grain measurements). However, evidence in the literature suggests that this may not always be the case - at least not with commonly used single-grain analysis procedures. Guérin et al., (2015) presented 19 samples with independent age control ranging between 2 and 46 ka and showed that the average ratio of the single-grain Central Age Model (CAM, Galbraith et al., 1999) age to the reference age was 0.90±0.02, indicating an average underestimation of ~10% by this standard frequentist approach. The word "frequentist" refers to frequentist statistics, in which probability is interpreted as the long run frequency of outcomes. By standard frequentist

approach, we refer here to using "*Analyst*" (Duller, 2015) to derive a set of paired D_e values and associated uncertainties which are then input into a dose estimation model such as CAM. Guérin et al. (2015) observed that the accuracy decreased with increasing age when standard frequentist analysis was applied, but when using a Bayesian approach ("BayLum", Combès et al., 2015; Combès and Philippe, 2017), the age control could be recovered. When using BayLum, the usual steps from dose response curve (DRC) analysis to age determination are combined into a single hierarchical model in which all parameters are sampled using Markov Chain Monte Carlo (MCMC) to produce posterior credible interval ranges (Combès et al., 2015; Combès and Philippe, 2017).

However, single-grain quartz analysis have also been found to be in agreement with independent age control (< 50 ka) using the standard frequentist approach (e.g., Jacobs et al., 2015; Jankowski et al., 2020; Colarossi et al., 2020; Kim et al., 2022), but only little (if any) testing of the accuracy of single-grain quartz OSL dating has been done for samples older than 50 ka, where a significant number of individual grains may not be able to accurately record the burial dose because of saturation effects. The purpose of this study is to investigate the accuracy of single-grain quartz OSL Single-Aliquot Regenerative (SAR, Murray and Wintle, 2000) dating for five OIS substage 5e samples with known absorbed doses in the 100-200 Gy range. Our primary questions are (i) how do the CAM and the Average Dose Model (ADM, Guérin et al., 2017) perform on these samples? (ii) Do commonly used rejection criteria improve equivalent dose estimation? (iii) Does application of BayLum (Christophe et al., 2023) result in more accurate quartz single grain age estimates than standard frequentist approaches?

2 Independent age control, samples and previous luminescence ages

For the purpose of testing single-grain OSL dating of quartz in the 100-200 Gy dose range, we chose five OIS substage 5e samples from two different sites (see below). Both sites have been determined to be early Eemian deposits by pollen analysis (Funder et al., 2002; Murray et al., 2007). The exact age of each stratigraphical position is not independently dated, but broadly said to have been deposited within a short period of time in the beginning of the Eemian. In our study, we adopt an early Eemian

age of 128 ka for all samples presented here and assign an uncertainty of 2 ka since the samples were likely deposited over a few thousand years.

Three samples (981007, 981009, 981013) are from Gammelmark, Denmark, which contains a sequence of marine sediments in upwards succession of a marine transgression layer. The samples were originally collected as part of an OSL dating application to test the accuracy of multi-grain OSL dating using sand-sized quartz grains (Murray and Funder, 2003). These authors defined an age for the whole marine section of between 133 and 125 ka (see section SI.1 for further information). Using quartz, these authors determined an average age of 119±2 ka (n=20) using the median as the dose estimator. Buylaert et al. (2011) revised these quartz ages (using early background subtraction, EBG, Ballarini et al., 2007, and an improved cosmic dose rate estimate) and derived an average age of 120±7 ka (n=20). These authors also presented fading corrected K-feldspar IRSL (IR_{50}) ages from which they obtained an average age of 128±3 ka (n=19). However, if the 8 most lower samples extracted from meltwater sand (where incomplete bleaching of the K-feldspar IR_{50} signal may be an issue) were rejected, an average age of 119±6 ka (n=11) was derived. Buylaert et al. (2012) used the pIRIR(50,290) protocol (Thiel et al., 2011) on four of the same K-rich extracts and obtained and average age of 112±4 ka (n=4). Table SI.1 provides an overview of the results of these previous luminescence ages for the specific samples directly relevant to this study.

The two remaining samples (H22547 and H22553) come from a section located on the Sula River, Russia (see Murray et al., 2007). They were sampled from a unit of marine sand characterized by fauna tied to the presence of warm coastal waters at the time of sediment deposition. By correlation with Western European pollen records (Funder et al., 2002), Murray et al. (2007) suggest a burial age of around 130 ka for the sediment column, while also stating that the entire sequence was deposited within a span of 5,000 years. Murray et al. (2007) determined an average SAR quartz multi-grain age of 108±2 ka (n=16) indicating an underestimate of ~15% of the expected age. However, using the SARA protocol (Mejdahl and Bøtter-Jensen, 1994), they obtained an average quartz age of 124±8 ka (n=16) consistent with the age control. Buylaert et al. (2008) measured K-rich feldspar for the same suite of samples using the fading corrected IR_{50} signal and obtained an average age of 107±2 ka

(n=16) and Buylaert et al. (2012) measured K-rich feldspar from three of the samples using the pIRIR(50,290) signal and obtained an average age of 121±7 ka (n=3). (see section SI.1 and Table SI.1 for further information).

These previously published pIRIR(50,290) ages are consistent with, or slightly younger than, the corresponding quartz ages. This is convincing evidence that the quartz at both these sites was bleached sufficiently at or before deposition to allow accurate dating (Murray et al., 2012, Möller and Murray, 2015).

3 Experimental details

3.1 Instrumentation

All quartz OSL measurements were made using Risø TL/OSL DA-20 readers (Bøtter-Jensen et al., 2010) equipped with Risø Single Grain Laser attachments (Bøtter-Jensen et al., 2003). For multi-grain measurements (Ø=8 mm), we stimulated quartz with blue LEDs (470 nm, ~80 mW/cm^2) and measured the resulting OSL signals through 7.5 mm Hoya U-340 detection filters. We loaded the samples onto stainless steel discs that were prepared with silicone oil in a circular area (Ø = 8 mm). For single-grain measurements, we stimulated individual quartz grains using a 10 mW Nd:YVO4 solid-state diode laser emitting at 532 nm. Individual grains were loaded onto special aluminium discs containing grain holes in a 10x10 array. We used two different hole-size discs, depending on the size fraction being measured. For a size fraction of 90-150 µm (samples 981007, -09 and -13), we used Ø=200 µm discs and for a size fraction of 180-250 µm, we used Ø=300 µm (samples H22547 and H22553). For beta irradiation we used calibrated ^{90}Sr/^{90}Y sources mounted on the readers (Bøtter-Jensen et al., 2010; Hansen et al., 2015; Autzen et al., 2022). The beta sources were calibrated in the same geometry as used for the measurements. The beta sources used in these experiments all have good spatial uniformity (i.e., <5% standard deviation across the sample area) and thus correcting for beta source non-homogeneity (e.g., using the approach developed by Lapp et al., 2012) would not result in significant changes to dose or scatter.

3.2 Dose rates and predicted doses

We adopt the radionuclide concentrations presented in Murray and Funder (2003) and Murray et al. (2007), which were all measured using high-resolution gamma spectrometry (Murray et al., 1987). We recalculated dose rates (see Table SI.2) using the conversion factors of Cresswell et al. (2018) and derived cosmic ray contributions from Prescott and Hutton (1994). Additionally, we assume an uncertainty of 2% for beta calibration and an internal quartz dose rate of 0.010±0.005 Gy/ka (consistent with Vanderberghe et al., 2008). We use the saturated water content for dose rate calculations as in the original publications (see section SI.2 for further details). Our newly derived dose rates for Gammelmark are larger than those originally published (by about 8%). This difference arises mainly from the conversion factors used. We expect the conversion factors of Cresswell et al. (2018) to generally result in larger dose rates, especially due to changes in ^{40}K factors. It may then seem curious that our estimated Sula dose rates are smaller compared to those originally published (about 6 and 14%). However, the decrease in dose rates for Sula is because we now use 0.01 Gy/ka as the internal dose rate coming from U and Th, whereas in Murray et al. (2007), a value of 0.06 Gy/ka (based on Mejdahl 1987) was assumed. Because of the low radionuclide concentrations for these Sula samples, the net effect is a decrease in total dose rate, despite the use of the conversion factors by Cresswell et al. (2018).

For an OSL age to match the independent age control, its equivalent dose should be equal to the age control multiplied by the environmental dose rate. This results in predicted doses of 186±8 Gy (981007), 188±9 Gy (981009), 217±10 Gy (981013), 132±7 Gy (H22547) and 95±6 5 Gy (H22553) (see Table SI.1). The uncertainties on the predicted doses have been derived from error propagation of the uncertainty assigned to the age control (±2 ka) and the total uncertainty on the environmental dose rate.

3.3 Measurement protocols and dose determination

For both multi-grain (MG) and single-grain (SG) measurements, we employed a Single-Aliquot Regenerative (SAR) protocol (Murray and Wintle, 2000) including at least one recycling, one IR depletion (for SG) and one recuperation measurement. We used a preheat of 260°C for 10 s, a cutheat

of 160°C and a test dose of 50 Gy. We stimulated MG aliquots for 40 s at 125 °C using the blue LEDs, and in signal analysis, we used the first 0.2 s of the measurement as the signal summation interval and the following 0.4 s as the background summation interval, i.e., early background subtraction, to maximise the fast-component contribution to the signal (Cunningham & Wallinga, 2010). However, we also used late background subtraction (LBG) with a background summation of the last 5 s. The average natural dose ratio of LBG to EBG is 0.94±0.03 (n=5 samples) where all individual sample ratios are consistent with unity (see section SI.7.2).

With single grains, we stimulated each grain for 1 s at 125 °C using the green laser. We used the first 0.05 s as the signal summation interval and the last 0.15 s as the background summation interval, i.e. late background subtraction (LBG). Applying EBG subtraction to SG data will result in the removal of a significant number of grains which is in fact fast-component dominated but appear to have significant medium/slow component because of the differences in effective stimulation power in the single grain measurement system (Thomsen et al., 2015). Thus, we do not consider it practical to apply EBG subtraction to our SG data.

Individual dose estimates were derived using "Analyst v4.56" (Duller, 2015). To fit single-grain dose response curves (DRCs), we used a single saturating exponential function forced through the origin (i.e., $L_x/T_x = A \times (1 - \exp(D/D_c))$, where A is the laboratory saturation level and D_c is a measure of the curvature of the dose response curve. To fit multi-grain DRCs, we used the sum of two saturating exponential functions forced through the origin to obtain a good fit to all regeneration points (see Fig. 1D and Table SI.5). However, we have also used a single saturating exponential to fit the MG DRCs, but often it was necessary to omit low dose regeneration point to obtain a reasonable fit to the DRCs in the region of interest. The average ratio of the dose estimated using the single saturating exponential fit to the double saturating exponential was 0.96±0.03 (n=5 samples), where all individual sample ratios also are consistent with unity (see section SI.7.3). We chose these fitting functions for the following reasons: for SG, a single saturating exponential function adequately captures the observed signal saturation without overfitting the data. For MG, the sum of two saturating exponentials appears better suited to account for late-saturating grains.

For BayLum analysis, we specified a saturating exponential function forced through the origin for both single-grain and multi-grain data. This means that we do not use the same fit to model multi-grain DRCs in BayLum as in the frequentist approach. This was a choice of necessity, as it is currently not possible to run BayLum using a double saturating exponential to fit the data. However, we note that when running BayLum using a saturation exponential and linear fit, results are indistinguishable from those obtained using a saturating exponential alone (BayLum EXP-to-BayLum EXP+LIN: 1.00±0.02, n=5). For most computations associated with BayLum, we made use of "Sophia" – a High-Performance-Computing-Cluster at DTU (Technical University of Denmark, 2019).

We define saturated aliquots as those whose natural sensitivity corrected OSL signal ($L_n/T_n \pm \sigma_n$) cannot be interpolated onto the corresponding DRC, where σ_n is the uncertainty assigned to L_n/T_n based on counting statistics, curve fitting errors and an instrument reproducibility of 0.5% and 2.5% per OSL measurement for multi-grain and single-grain measurements respectively (Thomsen et al., 2005).

3.4 Rejection criteria and dose models

For the multi-grain measurements, in the frequentist approach, we include all aliquots, which had a natural test dose response with an uncertainty known to better than 20% ("s_{Tn}<20%") and gave bounded dose estimates (i.e., aliquots that were not in saturation). However, we also investigated the effect of rejecting all aliquots which didn't have a recycling or IR depletion ratio within ±10% of unity and a recuperation value larger than 5% of the natural. Applying this additional rejection criteria has no effect on the average dose or the observed scatter (see section SI.7.1), i.e., the average dose ratio is 1.001±0.005 (n=5 samples) and the individual dose ratios are all consistent with unity. This is commonly observed (e.g., Murray et al., 2021 and references therein). In the Bayesian approach, we use all measured aliquots, including those in saturation.

For the single-grain measurements, we tested the application of three rejection criteria procedures: (i) include all grains for which the uncertainty on the natural test dose signal is less than 20 % ("s_{Tn}<20%"), (ii) include all grains where s_{Tn}<20%, the recycling ratio is within 2σ of unity, the IR depletion ratio is within 2σ of unity and the recuperation is less than 5% of the natural signal. Below

we refer to this set of criteria as the "standard" or "Std." rejection criteria, as these are commonly used in single-grain dating applications (e.g., Jacobs et al., 2008; Armitage et al., 2011; Demuro et al., 2019; Liu et al., 2022), (iii) include all grains which satisfy "s_{Tn}<20%" in combination with the D_c criterion (in previous publications referred to as the D_0 criterion, Thomsen et al., 2016). To determine an appropriate D_c threshold value, we plotted the apparent dose (calculated using either the CAM or the ADM) against increasing minimum accepted D_c. We then chose the D_c threshold to be the dose where the apparent central dose intersects the 1:1 line (e.g., Singh et al., 2017). This process is illustrated in the inset in Fig. 1H and 1I. Guo et al. (2017) suggested a revised version of the original D_c criterion in which the D_c threshold is increased until there are no grains in saturation. Although, this approach is likely to fall short if there are any grains for which the measured L_n/T_n values are significantly above the laboratory dose response curve (i.e. "Class 3" grains according to Yoshida et al., 2000). Nevertheless, we also tested this approach. We obtain a CAM dose ratio of 0.98±0.02 (n=5 samples) but exclude 180 additional grains in the Guo et al. (2017). As a result, the relative CAM dose error goes from between 5.4 and 5.9% (for the five samples) using our chosen D_c criterion, to ranging between 6.5 and 8.4%. In the frequentist single-grain approach, we exclude all grains not giving bounded dose estimates, but both include and exclude them in the Bayesian approach.

For multi-grain dose distributions, we compared the arithmetic central dose with the median dose and the central dose determined by BayLum. For single-grain data, we compared the following dose models: (i) CAM , (ii) ADM with intrinsic over-dispersions σ_m of 45% (Gammelmark) and 30% (Sula) Based on Guérin et al. (2017), these σ_m values have been determined from the relative over-dispersions determined in the dose recovery experiments using the s_{Tn}<20% criterion (see Table SI.4), and (iii) BayLum (assuming a log-normal dose distribution).

4 Luminescence characteristics

4.1 OSL stimulation curves and dose response curves

Fig. 1

Based on multi-grain measurements, our samples appear to be fast-component dominated (see Fig. 1), since the OSL stimulation curves of representative multi-grain Gammelmark and Sula samples are

very similar to those of Risø Calibration quartz; the latter of which is known to be fast-component dominated with no significant contribution from slower components (Jain et al., 2003; Hansen et al., 2015). Single-grain OSL stimulation curves are significantly more variable (see Fig.1B and 1C), but according to Thomsen et al. (2015) this is to be expected because the effective stimulation power is likely to vary from grain to grain and thus some fast-component dominated grains are likely to show stimulation curves which are slower than expected. For all Sula multi-grain aliquots, the average recycling ratio is 1.03±0.02, the IR depletion ratio is 0.952±0.011 and the average recuperation is −0.03±0.14% of the natural (n = 64). For Gammelmark, the corresponding numbers are 1.004±0.005, 0.973±0.005 and 0.14±0.02% of the natural signal (n = 94). The median D_c values for the double saturating exponential fits of the DRCs are 25 Gy and 185 Gy (n = 154). Typical multi-grain dose response curves are shown in Fig. 1D. Fig. 1E and 1F shows representative single-grain OSL dose response curves for Gammelmark (981009) and Sula (H22553) respectively, highlighting the usual considerable between-grain variability for single-grain measurements (e.g., Duller, 2008). Fig. 1H and 1I shows the D_c distribution at the single-grain level for 981009 (predicted dose of 188 Gy) and H22553 (predicted dose of 95 Gy) respectively. We observe a wide dispersion of values, with a great many below 100 Gy. The application of the D_c criterion is shown in the inset, where increasing the D_c threshold for accepting grains into analysis creates an initial increase in the derived dose, after which the curve plateaus. Based on the above, we consider both Gammelmark and Sula quartz suitable for OSL dating.

4.2 Recovery of laboratory given doses

Fig. 2

Laboratory dose recovery experiments (both multi-grain and single-grain) were undertaken by bleaching aliquots, either using the blue LEDs twice for 100 s (with an intervening pause of 10 ks) or in a daylight simulator for >2 h before a known laboratory dose was given and measured. No consistent difference could be seen between the different modes of bleaching and thus dose recovery results have been combined (see Fig. SI.3.1 for further details). Thermal transfer was assessed by measuring the residual dose in a subset of aliquots after blue LED bleaching. Residual doses of

0.49±0.09 Gy (n=12, H22553) and 0.65±0.07 Gy (n=12, 981013) was measured and thus thermal transfer is considered small for our samples. The ability of SAR to recover doses administered in the laboratory does not necessarily convey its ability to recover natural doses, but it is the most complete test of protocol performance (Murray et al., 2021). By convention, we deem dose recovery results within 10% of unity as acceptable (Wintle and Murray, 2006).

The results of the multi-grain quartz OSL dose recovery experiments are shown in Fig. 2. Our multi-grain SAR protocol recovered doses given to the Gammelmark samples over the full range of tested doses (25-300 Gy), with an overall average dose recovery ratio of 1.00±0.02 (n=9). For the Sula samples, the corresponding average ratio for given doses ranging between 25 and 388 Gy, is 1.23±0.04 (n=9); possibly indicating that there is uncorrected sensitivity change in the first measurement cycle. In the first publication on multi-grain quartz OSL for these samples (Murray et al., 2007), it was observed that SAR, on average, underestimated the age control by ~15% (n=16), but that the SARA (single aliquot regeneration added dose) protocol (Mejdahl and Bøtter-Jensen, 1994) on average recovered the age control (ratio of 0.95±0.06 to age control). We have applied the SARA protocol to two samples from Gammelmark and one from Sula and find that all slopes are consistent with unity (average of 1.00±0.02, n=3, see Fig. SI.3.2) and thus the unacceptable SAR dose recovery ratios for the Sula samples are not caused by first cycle sensitivity changes. It is interesting to note that the dose recovery ratios for H22553 are only acceptable for the very high given doses > 300 Gy.

In the single grain dose recovery experiments, the given dose range between 25 and 250 Gy (981009 and H22553). The results are shown in Fig. 2A (981009), Fig. 2B (H22553) and in Table SI.4. Applying the standard rejection criteria does not have a significant effect on the estimated dose or the dose over-dispersion (OD) for any of the samples, regardless of the dose estimation model used. Using CAM, the average dose ratio of ”Std.” to ”s_{Tn}<20%” is 0.985±0.014 (n=9) and for BayLum, the ”Std.” to ”s_{Tn}<20%” ratio is 0.99±0.02 (n=9). In terms of OD, the same rejection criteria comparison yields a ratio of 1.00±0.03 (n=9). Thus, the only effect of applying the ”Std.” rejection criteria is to reduce the grain population by about 30%. It is worth noting that, when applying the D_c criterion, the dose estimate increases smoothly as a function of given dose, i.e., the ratio of CAM_{Dc} to $CAM_{sTn<20\%}$

increases from 1.07±0.06 at 25 Gy to 1.66±0.18 (at 250 Gy). Thus, application of the D_c criterion increases the recovered dose, but simultaneously decreases the observed OD (see Table SI2). The ratio of OD_{Dc} to $OD_{sTn<20\%}$ decreases continuously from 1.00±0.14 (at 50 Gy) to 0.49±0.16 % (at 250 Gy). This ratio is poorly known at 25 Gy (i.e., 0.76±0.26) because relatively few grains were measured at this dose. Application of the D_c criterion also reduces the number of saturated grains to <6%. Interestingly, the dose estimated by BayLum is relatively insensitive to the application of the D_c criterion (ratio of 0.974±0.013, n=9 samples). A similar effect in dose recovery experiments was previously observed by Heydari and Guérin (2018). As expected, the OD increases as a function of given dose (Thomsen et al., 2012) from 19.8±1.0 % (at 25 Gy) to 51±5% (at 250 Gy). Applying the D_c criterion gives OD_{Dc} values for all given doses of ~ 25 %. Thus, applying the D_c criterion reduces the apparent intrinsic over-dispersion and makes it independent of dose.

Fig. 2A summarises the dose recovery results for 981009 (Gammelmark). A residual dose of 1.2±0.4 Gy (n = 21, calculated using the un-logged version of CAM, Arnold et al., 2009) was measured, but not subtracted from the dose measured in the dose recovery experiments. The CAM dose recovery ratio decreases with given dose. At a given dose of 70 Gy, the ratio is 1.08±0.02 (n=278) and it decreases to 0.68±0.04 (n = 106) at 250 Gy. The CAM dose recovery is only acceptable for given doses ranging between 70 and 128 Gy. When the D_c criterion is applied, the same decreasing trend in the absolute dose recovery ratio is observed, but now acceptable dose recoveries (dose recovery ratio within 10% of unity) are in the range between 100 and 250 Gy. Hence, in a laboratory prepared sample, we are able to recover high given doses, even when the majority of grains do not give bounded dose estimates (65% at 250 Gy). However, the application of the D_c criterion in these samples, particularly at high doses, is very expensive in terms of grain-loss (and therefore also precision). For BayLum, the Bayesian approach allows inclusion of saturated grains. If we include all grains with s_{Tn}<20% in BayLum, we successfully recover the given dose in the interval 128-250 Gy (average dose recovery ratio of 1.087±0.013, n=3). To test the influence of including saturated grains, we also ran BayLum without including these saturated grains ("$BayLum_{no\ sat}$"). We observe acceptable dose recovery ratios in the interval 100-250 Gy (average dose recovery ratio of 0.99±0.03, n=4). Thus,

in these dose recovery experiments, it appears that BayLum successfully recover the given dose in the 128-250 dose range regardless of whether saturated grains are included or not. This is surprising but probably testifies to the power of the Bayesian approach to dose estimation.

Fig. 2B summarises the dose recovery results for H22553 (Sula). A residual dose of 1.2±0.3 Gy (n = 33, calculated using the un-logged version of CAM, Arnold et al., 2009) was measured, but not subtracted from the dose measured in the dose recovery experiments. In general, we observe the same pattern as for the Gammelmark sample. We have unacceptable dose recovery ratios at low doses (25 and 50 Gy) but acceptable dose recoveries for the two other investigated doses for this sample (i.e., 85 and 110 Gy).

5 Natural equivalent doses

5.1 Multi-grain equivalent doses

Table 1 summarizes our multi-grain quartz results. The equivalent doses have been normalized to the predicted dose. Multi-grain dose distributions have relative standard deviations ranging between 20 and 38 % (Fig. SI.5) and appear positively skewed.

Table 1

This is often interpreted as a sign of significant incomplete bleaching (e.g., Mellett et al., 2012; Alexanderson and Bernhardson, 2016; Perilla-Castillo et al., 2023). However, these data were obtained using multi-grain aliquots, each containing thousands of grains. It is therefore to be expected that averaging effects would prevent the use of the shape of the dose distribution to detect incomplete bleaching. Also, significant incomplete bleaching of these samples can confidently be ruled out because of the relatively good agreement between K-rich feldspar IRSL ages and quartz ages (Murray and Funder, 2003; Murray et al., 2007; Buylaert et al., 2008, 2011, 2012) despite the vastly different bleaching rates of the two dosimeters (e.g., Murray et al., 2012). However, it is interesting to note that for three of the five samples (981009, 981013, H22553), the natural sensitivity corrected L_n/T_n values appear to be normally distributed (data not shown). This implies that for these samples, it is the interpolation onto the curving part of the laboratory constructed DRCs which is causing the observed skewness. In this case, it can be argued that the median dose is more accurate than the average dose

(Murray and Funder, 2003; Murray et al., 2021). However, for these samples there is no significant difference between the average and median natural doses. These are also indistinguishable from the BayLum results (Table 1).

In the SARA protocol, aliquots containing their natural doses are given increasing additional beta doses to determine the relationship between the measured dose and the added dose. The purpose of the SARA is to derive an equivalent dose while accounting for natural-cycle sensitivity changes which may occur in quartz when first preheated and measured. In Fig. SI.3.2, we show the SARA results for three samples (981007, 981013 and H22553). The average ratio between our resulting SARA D_e values and the expected doses is 1.03±0.05 (n=3).

Fig. 3

Thus, using multi-grain aliquots we can recover the expected dose of the independent age control using both SAR and SARA. So at least in the case of Sula, it appears that inferences made about the performance of SAR using multi-grain dose recovery experiments are of limited value.

Table 2

5.2 Single-grain equivalent doses

We measured the natural OSL signals from a total of 34,700 individual grains from the five samples investigated here. Of these grains, 5.3% gave an acceptable test dose response (i.e., s_{Tn}<20%). However, only about 60% of these grains gave bounded dose estimates, i.e., ~40% of the light-giving grains appeared to be in saturation. Table 2 show these characteristics for the individual samples. All five samples have equivalent dose distributions characterized by positive skewness and over-dispersions of more than 40% (see Fig. SI.6 and Table 2). All dose distributions contain a wide range of dose estimates. For instance, in sample H22547, estimates range from 1±8 Gy to 457±169 Gy, while sample 981007 estimates range from 1±2 Gy to 457±141 Gy. We observe the presence of few grains with doses consistent with zero (average: 3.7±0.2% of grains per sample, n = 5). These grains are unlikely to arise from post-depositional mixing at these depths (>3 m) and all grain holes in the single-grain discs were screened for contamination (i.e., stuck grains) prior to use. Laboratory

contamination can never be completely ruled out, but we do not consider it to be a likely explanation. We here note that reporting on apparent zero-dose grains in old samples is not unusual (e.g., Arnold and Roberts, 2011; Singh et al., 2017) and in the absence of convincing external reasons (e.g., post-depositional mixing) for rejecting these grains, we must accept that these outliers are simply an indication of the scatter in these measurements.

When rejecting grains due only to the s_{Tn}<20% criterion, we see from Table 2 that for the Gammelmark samples (9810xx), about 45% of grains do not give bounded dose estimates, i.e., they are lost from frequentist analysis due to saturation. This is of great concern, because it must mean that the grains for which a bounded dose estimate could be derived are likely to underestimate the true burial age (e.g., Thomsen et al., 2016). Additionally, for the Gammelmark samples, we also observe large over-dispersion values, with an average value of 62.7±0.7% (n = 3). In the literature, it has been suggested that single-grain over-dispersion values larger than 20% could suggest significant incomplete bleaching (e.g., Olley et al., 2004; Arnold et al., 2009). However, this can confidently be ruled out here, because of the previously published good agreement with feldspar ages (Buylaert et al., 2008, 2011, 2012). In addition, the intrinsic overdispersion (derived from dose recovery experiments) are significantly larger than this threshold value; they range between 46 ± 4% and 51 ± 5% for given doses of 180 Gy and 250 Gy (Table SI.4). The Sula samples have fewer grains in saturation (~35 %) and smaller over-dispersion values (i.e., 58±3% and 42±4% for H22547 and H22553 respectively). The difference in over-dispersion between the two Sula samples could imply that a significant portion of the over-dispersion is in fact due to the curvature of the region of DRC-interpolation and not incomplete bleaching, since the sample with the lower over-dispersion is also the sample with a lower predicted dose (95±6 Gy vs 132±7 Gy).

Applying the “Std.” rejection criteria (recycling ratio, OSL IR depletion ratio and recuperation) does not make a significant difference in terms of central dose, overdispersion or proportion of grains in saturation from accepting all grains giving a detectable natural test dose signal (here defined as s_{Tn}<20%). While examples exist of “Std.” rejection criteria having an effect (e.g., Jakobs et al., 2006), observations like ours have been reported numerous times in the literature (e.g., Thomsen et al., 2012;

Geach et al., 2015; Guérin et al., 2015a; Kristensen et al., 2015; Thomsen et al., 2016; Guérin et al., 2017; Singh et al., 2017; Murray et al., 2021; Marquet et al., 2023). Only with the D_c criterion are we able to reduce the relative number of grains in saturation, i.e., the average saturation of 40±2 % is reduced to 9±2 %. The over-dispersion is also reduced, from an average relative OD of 58±4 % to 41±3 %.

In the following, we focus only on results from application of the s_{Tn}<20% criterion, unless otherwise specified. Generally, CAM doses are significantly smaller than ADM doses for these five samples (average CAM-to-ADM ratio is 0.914±0.014, n=5). However, both CAM and ADM are consistently and significantly smaller than the BayLum doses (average ADM-to-BayLum ratio is 0.63±0.04, n=5). For BayLum, it makes a significant difference if grains which give no bounded estimates in frequentist analysis are included or not. The average ratio between BayLum and BayLum only giving bounded dose estimates ($BayLum_{no\ sat}$) is 1.29±0.05 (n=5). This contrasts with what we observed in the dose recovery tests, where the BayLum results were largely unaffected by the inclusion or omission of the saturated grains. Applying the D_c criterion decreases the difference between ADM and BayLum by ~20% (average ADM_{Dc}-to-$BayLum_{Dc}$ ratio is 0.79±0.03, n=5), partly because ADM estimates increase (by 28±7 %, n=5) and partly because BayLum estimates decrease (by 6.0±1.1 %, n=5).

5.3 Comparisons with predicted doses from the independent age control

Fig. 4

Fig. 3 shows the multi-grain results obtained using the arithmetic average of the multi-grain dose results. We observe that four of five multi-grain results are consistent with unity within 2σ (MG doses to Predicted doses range between 0.84±0.05 and 1.15±0.10). Only the highest dose sample (981013) underestimates the age control (measured MG dose to predicted dose of 0.84±0.05). For the Gammelmark samples, the average ratio between multi-grain SAR equivalent doses presented here and multi-grain SAR equivalent doses published by Murray and Funder (2003) is 1.05 ± 0.03 (n = 3) consistent within 2σ. However, for the Sula samples, our doses are consistently larger than those of Murray et al. (2007). The average ratio between SAR equivalent doses presented here and those

published by Murray et al. (2007) is 1.38±0.16 (n = 2). It may be that corrections to the dose absorbed by calibration quartz would increase past equivalent doses by up to 8% (Autzen et al., 2022), but this increase is incidentally almost accounted for using updated conversion factors (Cresswell et al., 2018) and the assumed internal dose rate in quartz (based on Vanderberghe et al., 2008).

With respect to the single-grain measurements, Fig. 3B clearly shows that both CAM and AMD dose models severely underestimate the age control for four of the five samples with ~45% and ~40%, respectively. Only for the sample with the lowest dose (H22553 with a predicted dose of 95±5 Gy) can we recover the predicted dose within 3σ (CAM and ADM ratios of 0.84±0.06 and 0.87±0.06, respectively). When we apply the D_c criterion (Fig. 3B) the average ADM_{Dc}-to-predicted dose ratio improves to 0.75±0.06 (for H22553 it is 0.94±0.07). However, except for H22553, this ratio is still unacceptably small. BayLum-to-predicted dose ratios range from 0.75±0.07 to 1.14±0.08, with only the Sula samples consistent with unity, but only when saturated grains are included. Without these grains, the average ratio is 0.74±0.07. If we used BayLum on the D_c filtered data sets, the ratios are on average decreased by 6.0±1.1% in contrast to the dose recovery experiments where application of the D_c criterion did not result in a significant difference.

In Fig. 4 we plot the ratios of measured equivalent doses and predicted doses as a function of the relative number of the light giving grains appearing to be in saturation (i.e., n_{sat}). For all dose models there appears to be a correlation between the proportion of saturated grains and how well the predicted dose can be recovered.

BayLum is most successful in recovering the predicted dose for the two Sula samples (H22547 and H22553), which also have the smallest proportion of saturated grains. For the three Gammelmark samples (981007, -09 and -13), not even BayLum is able to recover the predicted dose, and there is a clear correlation between the recovery ratio and the relative number of saturated grains.

However, why is it that we recover the age control using multi-grain aliquots and performs so poorly using single grain measurements? If we sum our single grain data (before analysing the data), do we then obtain the same results as we do using true multi-grain measurements? To investigate this, we

have made "synthetic" multi-grain aliquots from our single-grain data (using the "*Sum all grains*" function in Analyst) and processed those as the true multi-grain data (see Table SI.7.4). Only for sample H22553 (the sample with the lowest dose) are the synthetic aliquot doses indistinguishable from the true multi-grain measurements. For the remaining samples, the synthetic aliquots underestimate the true multi-grain doses by 21±3%.

6 Discussion

Had we a-priori chosen to use quartz single-grain OSL dating using the standard procedures outlined herein, we would have severely underestimated the burial ages. For example, the mean $ADM_{Std.}$-to-predicted dose ratio is 0.57±0.09 (n=5). If we had been concerned with the level of saturation, and therefore applied the D_c criterion in order to remove grains not able to record the absorbed dose (but giving bounded dose estimates nonetheless), the number of saturated grains would be reduced to between 3 and 12 % of grains, but we would still significantly underestimate the ages using either CAM or ADM (ADM_{Dc}-to-predicted dose ratio: 0.75±0.06, n=5). Had we instead chosen to date these samples using both conventional single-grain and multi-grain procedures, we would have been left with contrasting results, i.e., the average $CAM_{Std.}$-to-multi-grain ratio is 0.57±0.06 (n=5). Given no external source of support, such as independent age control, we might favour single-grain results since: "*... the OSL characteristics of each grain have been individually evaluated against objective quality assurance criteria, and only grains considered reliable contribute to the final burial dose estimate*" (Arnold et al., 2013). This would clearly not be a suitable choice for these samples. Had we a-priori chosen the standard multi-grain approach, we would have obtained accurate ages, unaware that our dose estimates derive from averaging effects of a large fraction of grains whose signals do not interpolate onto their individual dose response curves. To explain why our more conventional single-grain procedures underestimate the age control so dramatically, one could argue that we should not at all be surprised given that 30-45% of the light giving grains are in apparent dose saturation for these samples. If saturated grains more likely belong to a higher-than-median dose sub-population, having to take these grains out of analysis would bias our estimates toward lower doses (Murray and Funder, 2003). But of course, we still significantly underestimate the age control when applying the D_c

criterion, which aims at eliminating saturation effects. Even if we are able to account for why single-grain procedures perform so poorly for these samples, we are still left with the question of why our multi-grain procedure performs so well.

From basic principles, one would expect that the sum of the OSL signals from individual grains would equal that of multi-grain, if we assume that there is no significant effect of using different stimulation wavelengths and different stimulation powers. It has previously been shown that there is no significant difference in DRC shape between using blue (MG dose estimation) and green (SG dose estimation) light stimulation, but that the high power green laser stimulation used in the single grain system gives a significant reduction in the average D_c value of 23±3% (Thomsen et al., 2015). Thus, one should not expect the sum of the single grain OSL data to be equal to that of multi-grain data because of this difference. In addition, Thomsen et al. (2015) also showed that the large variability observed in the shape of quartz single grain stimulation curves is largely caused by varying effective stimulation power, so summing individual single grain data to produce synthetic small multi-grain aliquots should not necessarily be expected to be equivalent to true multi-grain measurements. Hence it is not surprising that the synthetic multi-grain aliquots are significantly different from the true multi-grain data (except for sample H22553).

If we apply BayLum to our single-grain data, then we are able to recover the expect dose on average. Using either only the s_{Tn}<20% criterion (average BayLum-to-age control dose ratio of 0.95±0.08, n=5) or the commonly used rejection criteria ("Std.") (average BayLum-to-age control dose ratio of 0.96±0.08, n=5), we recover the age control within 1σ - but only when saturated grains are also included in the BayLum model. If left out of the analysis, BayLum underestimates doses in a similar fashion to the D_c criterion with ADM. However, it should be noted that for samples with more than 40% of the grains in saturation, BayLum is also having trouble recovering the expected dose, i.e., the ratio is 0.82±0.06 (n=3 samples). Given that the most accurate doses are obtained using standard multi-grain analysis or single-grain measurements analysed with BayLum (including saturated grains), is it then the case that saturated grains provide information necessary for accurate burial dose estimation at high doses (where incomplete bleaching can confidently be ruled out)? One contrary

argument is that we should expect larger net dose estimates when saturated grains are included in the analysis; for samples outside the dating range, this will create estimates closer to the true age by chance. Of course, it is then curious that our estimates, on average, match the predicted doses and do so over a range of 100 Gy. In contrast, in 4 out of 5 samples, conventional single-grain procedures significantly underestimate in this dose range (the exception is sample H22553, with a predicted dose of 95±6 Gy).

7 Conclusions

In this study we investigate the accuracy of quartz OSL multi-grain and single-grain techniques on five previously published samples of known age (128±2 ka) – three from Gammelmark, Denmark, and two from Sula, Russia. For the single-grain analysis, we compare three sets of acceptance criteria: (i) all grains with a relative uncertainty on the natural test dose response (s_{Tn}<20%) is included, (ii) only grains with s_{Tn}<20% and a set of "standard" acceptance criteria are included, i.e., both the recycling ratio and the IR depletion ratio are within 2σ of unity, and the recuperation (measured after the largest regeneration dose) is less than 5% of the natural signal, (iii) grains with s_{Tn}<20% and with a D_c value larger than a certain threshold value are included. We compare equivalent doses to the predicted doses from the independent age control using three different dose models: ADM, CAM and BayLum. For our samples, we find that standard multi-grain measurements on average recover the age control, but that standard single-grain analysis using CAM severely underestimates the age control of all samples, i.e., the CAM-to-Predicted dose ratios range between 0.43±0.03 and 0.84±0.07 (n=5). Reducing the number of grains likely to be affected by saturation effects through the application of the D_c criterion improves the accuracy, but the estimated doses still significantly underestimate the age control (ratios range between 0.63±0.05 and 0.94±0.08), except for the sample with the smallest absorbed dose (H22553) which is consistent with the age control. Using BayLum, the dose estimate is consistent with age control for only two of the five samples.

On the other hand, four of five MG measurements come within 2σ of predicted values. We conclude that for these samples, multi-grain OSL dating of quartz is the simplest and most accurate

chronometer. Our results may have considerable implications for the reliability of published single-grain OSL dating of quartz in the 100-200 Gy natural dose range.

Acknowledgements: This work was supported by the Independent Research Fund Denmark [grant number 9040-00308B].

References

Alexanderson, H. and Bernhardson, M. (2016). OSL dating and luminescence characteristics of aeolian deposits and their source material in Dalarna, central Sweden. Boreas, 45(4):876–893.

Armitage, S. J., Jasim, S. A., Marks, A. E., Parker, A. G., Usik, V. I., and Uerpmann, H.-P. (2011). The Southern Route Out of Africa: Evidence for an Early Expansion of Modern Humans into Arabia. Science, 331(6016):453–456.

Arnold, L., Roberts, R., Galbraith, R., and DeLong, S. (2009). A revised burial dose estimation procedure for optical dating of young and modern-age sediments. Quaternary Geochronology, 4(4):306–325.

Arnold, L. J. and Roberts, R. G. (2011). Paper I – Optically Stimulated Luminescence (OSL) dating of perennially frozen deposits in north-central Siberia: OSL characteristics of quartz grains and methodological considerations regarding their suitability for dating. Boreas, 40(3):389–416.

Arnold, L. J., Demuro, M., Navazo, M., Benito-Calvo, A., and Pérez-González, A. (2013). OSL dating of the Middle Palaeolithic Hotel California site, Sierra de Atapuerca, north-central Spain. Boreas, 42(2):285–305.

Autzen, M., Andersen, C., Bailey, M., and Murray, A. (2022). Calibration quartz: An update on dose calculations for luminescence dating. Radiation Measurements, 157:106828.

Ballarini, M., Wallinga, J., Wintle, A., and Bos, A. (2007). A modified SAR protocol for optical dating of individual grains from young quartz samples. Radiation Measurements, 42(3):360–369.

Buylaert, J., Murray, A., and Huot, S. (2008). Optical dating of an Eemian site in Northern Russia using K-feldspar. Radiation Measurements, 43(2):715–720.

Buylaert, J.-P., Huot, S., Murray, A. S., and Van Den Haute, P. (2011). Infrared stimulated luminescence dating of an Eemian (MIS 5e) site in Denmark using K-feldspar. Boreas, 40(1):46–56.

Buylaert, J.-P., Jain, M., Murray, A. S., Thomsen, K. J., Thiel, C., and Sohbati,R. (2012). A robust feldspar luminescence dating method for Middle and Late Pleistocene sediments. Boreas, 41(3):435–451.

Bøtter-Jensen, L., Andersen, C., Duller, G., and Murray, A. (2003). Developments in radiation, stimulation and observation facilities in luminescence measurements. Radiation Measurements, 37(4):535–541.

Bøtter-Jensen, L., Thomsen, K. J., and Jain, M. (2010). Review of optically stimulated luminescence (OSL) instrumental developments for retrospective dosimetry. Radiation Measurements, 45:253–257.

Christophe, C., Philippe, A., Guérin, G., Mercier, N., and Guibert, P. (2018). Bayesian approach to OSL dating of poorly bleached sediment samples: Mixture Distribution Models for Dose (MD2). Radiation Measurements, 108:59–73.

Christophe, C., Philippe, A., Kreutzer, S., Guérin, G., and Baumgarten, F. (2023). BayLum: Chronological Bayesian Models Integrating Optically Stimulated Luminescence and Radiocarbon Age Dating. https://cran.r-project.org/package=BayLum.R package version 0.3.1.

Colarossi, D., Duller, G., Roberts, H., Tooth, S., and Botha, G. (2020). A comparison of multiple luminescence chronometers at Voordrag, South Africa. Quaternary Geochronology, 60:101094.

Combès, B. and Philippe, A. (2017). Bayesian analysis of individual and systematic multiplicative errors for estimating ages with stratigraphic constraints in optically stimulated luminescence dating. Quaternary Geochronology, 39:24–34.

Combès, B., Philippe, A., Lanos, P., Mercier, N., Tribolo, C., Guérin, G., Guibert, P., and Lahaye, C. (2015). A Bayesian central equivalent dose model for optically stimulated luminescence dating. Quaternary Geochronology.

Cunningham, A. C., & Wallinga, J. (2010). Selection of integration time intervals for quartz OSL decay curves. Quaternary Geochronology, 5(6), 657-666.

Cresswell, A., Carter, J., and Sanderson, D. (2018). Dose rate conversion parameters: Assessment of nuclear data. Radiation Measurements, 120:195–201.

Demuro, M., Arnold, L. J., Aranburu, A., Gómez-Olivencia, A., and Arsuaga, J.-L. (2019). Single-grain OSL dating of the Middle Palaeolithic site of Galería de las Estatuas, Atapuerca (Burgos, Spain). Quaternary Geochronology, 49:254–261.

Demuro, M., Arnold, L. J., González-Urquijo, J., Lazuen, T., and Frochoso, M. (2023). Chronological constraint of Neanderthal cultural and environmental changes in southwestern Europe: MIS 5–MIS 3 dating of the Axlor site (Biscay, Spain). Journal of Quaternary Science, 38(6):891–920.

Duller, G. A. T. (1994). Luminescence dating of poorly bleached sediments from Scotland. Quaternary Science Reviews, 13:521–524.

Duller, G. A. (2008). Single-grain optical dating of quaternary sediments: why aliquot size matters in luminescence dating. Boreas, 37(4):589–612.

Duller, G. (2015). Analyst software package for luminescence data: overview and recent improvements. Ancient TL, 33:35 – 42.

Fu, X., Cohen, T. J., and Fryirs, K. (2019). Single-grain OSL dating of fluvial terraces in the upper Hunter catchment, southeastern Australia. Quaternary Geochronology, 49:115–122.

Funder, S., Demidov, I., and Yelovicheva, Y. (2002). Hydrography and mollusc faunas of the Baltic and the White Sea–North Sea seaway in the Eemian. Palaeogeography, Palaeoclimatology, Palaeoecology, 184(3):275–304.

Galbraith, R. F., Roberts, R. G., Laslett, G. M., Yoshida, H., and Olley, J. M. (1999). Optical dating of single and multiple grains of quartz from Jinmium rock shelter, northern Australia: Part I, experimental design and statistical models. Archaeometry, 41(2):339–364.

Geach, M., Thomsen, K., Buylaert, J.-P., Murray, A., Mather, A., Telfer, M., and Stokes, M. (2015). Single-grain and multi-grain OSL dating of river terrace sediments in the Tabernas Basin, SE Spain. Quaternary Geochronology, 30:213–218.

Guérin, G., Christophe, C., Philippe, A., Murray, A., Thomsen, K., Tribolo, C., Urbanova, P., Jain, M., Guibert, P., Mercier, N., Kreutzer, S., and Lahaye, C. (2017). Absorbed dose, equivalent dose, measured dose rates, and implications for OSL age estimates: Introducing the Average Dose Model. Quaternary Geochronology, 41:163–173.

Guérin, G., Combès, B., Lahaye, C., Thomsen, K. J., Tribolo, C., Urbanova, P., Guibert, P., Mercier, N., and Valladas, H. (2015). Testing the accuracy of a Bayesian central-dose model for single-grain OSL, using known-age samples. Radiation Measurements, 81:62–70.

Hansen, V., Murray, A. S., Buylaert, J.-P., Yeo, E. Y., and Thomsen, K. J. (2015). A new irradiated quartz for beta source calibration. Radiation Measurements, 81:123– 127.

Heydari, M. and Guérin, G. (2018). OSL signal saturation and dose rate variability: Investigating the behaviour of different statistical models. Radiation Measurements, 120:96–103. 15th International Conference on Luminescence and Electron Spin Resonance Dating, 11-15 September 2017.

Jacobs, Z., Hayes, E. H., Roberts, R. G., Galbraith, R. F., and Henshilwood, C. S. (2013). An improved OSL chronology for the Still Bay layers at Blombos Cave, South Africa: further tests of single-grain dating procedures and a re-evaluation of the timing of the Still Bay industry across southern Africa. Journal of Archaeological Science, 40(1):579–594.

Jacobs, Z., Li, B., Jankowski, N., and Soressi, M. (2015). Testing of a single grain OSL chronology across the Middle to Upper Palaeolithic transition at Les Cottés (France). Journal of Archaeological Science, 54:110–122.

Jacobs, Z. and Roberts, R. G. (2015). An improved single grain OSL chronology for the sedimentary deposits from Diepkloof Rockshelter, Western Cape, South Africa. Journal of Archaeological Science, 63:175–192.

Jacobs, Z., Roberts, R. G., Galbraith, R. F., Deacon, H. J., Grün, R., Mackay, A., Mitchell, P., Vogelsang, R., and Wadley, L. (2008). Ages for the Middle Stone Age of Southern Africa: Implications for Human Behavior and Dispersal. Science, 322(5902):733–735.

Jacobs, Z., Duller, G. A. T., Wintle, A. G. (2006). Interpretation of single grain De distributions and calculation of De. Radiation Measurements, 41(3): 264-277.

Jain, M., Murray, A., and Bøtter-Jensen, L. (2003). Characterisation of blue-light stimulated luminescence components in different quartz samples: implications for dose measurement. Radiation Measurements, 37(4-5):441–449.

Jankowski, N. R., Stern, N., Lachlan, T. J., and Jacobs, Z. (2020). A high-resolution late Quaternary depositional history and chronology for the southern portion of the Lake Mungo lunette, semi-arid Australia. Quaternary Science Reviews, 233:106224.

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.

Kim, S.-J., Choi, J.-H., Lim, H. S., Shin, S., Yeo, E.-Y., Weon, H.-J., and Heo, S. (2022). Multiple and single grain quartz OSL dating of dolmens in Jungdo, central Korean Peninsula. Geosciences Journal, 26:487–498.

Kristensen, J. A., Thomsen, K. J., Murray, A. S., Buylaert, J.-P., Jain, M., and Breuning-Madsen, H. (2015). Quantification of termite bioturbation in a savannah ecosystem: Application of OSL dating. Quaternary Geochronology, 30:334–341.

Liu, R., Nian, X., Zhang, W., Qiu, F., Wang, Z., Lin, Q., Shu, J., and Liu, N. (2022). Luminescence dating of the late Quaternary sediments in Hangzhou Bay, China. Quaternary Geochronology, 70:101302.

Marquet, J.-C., Freiesleben, T. H., Thomsen, K. J., Murray, A. S., Calligaro, M., Macaire, J.-J., Robert, E., Lorblanchet, M., Aubry, T., Bayle, G., et al. (2023). The earliest unambiguous Neanderthal engravings on cave walls: La Roche-Cotard, Loire Valley, France. Plos one, 18(6):e0286568.

Mejdahl, V. (1987). Internal radioactivity in quartz and feldspar grains. Ancient TL, 5, 10-17.

Mejdahl, V. and Bøtter-Jensen, L. (1994). Luminescence dating of archaeological materials using a new technique based on single aliquot measurements. Quaternary Science Reviews, 13(5-7):551–554.

Mellett, C. L., Mauz, B., Plater, A. J., Hodgson, D. M., and Lang, A. (2012). Optical dating of drowned landscapes: A case study from the English channel. Quaternary Geochronology, 10:201–208.

Murray, A., Arnold, L. J., Buylaert, J.-P., Guérin, G., Qin, J., Singhvi, A. K., Smedley, R., and Thomsen, K. J. (2021). Optically stimulated luminescence dating using quartz. Nature Reviews Methods Primers, 1(1):72.

Murray, A. and Funder, S. (2003). Optically stimulated luminescence dating of a Danish Eemian coastal marine deposit: a test of accuracy. Quaternary Science Reviews, 22(10):1177–1183.

Murray, A., Marten, R., Johnston, A., and Martin, P. (1987). Analysis for naturally occurring radionuclides at environmental concentrations by gamma spectrometry. Journal of Radioanalytical and Nuclear Chemistry, 115(2):263 – 288.

Murray, A., Svendsen, J., Mangerud, J., and Astakhov, V. (2007). Testing the accuracy of quartz OSL dating using a known-age Eemian site on the river Sula, northern Russia. Quaternary Geochronology, 2(1):102–109.

Murray, A., Thomsen, K., Masuda, N., Buylaert, J., and Jain, M. (2012). Identifying well-bleached quartz using the different bleaching rates of quartz and feldspar luminescence signals. Radiation Measurements, 47(9):688–695.

Murray, A. and Wintle, A. (2000). Luminescence dating of quartz using an improved single-aliquot regenerative-dose protocol. Radiation Measurements, 32(1):57–73.

Möller, P. and Murray, A., (2015). Drumlinised glaciofluvial and glaciolacustrine sediments on the Småland peneplain, South Sweden – new information on the growth and decay history of the Fennoscandian Ice Sheets during MIS 3. Quat. Sci. Rev. 122: 1-29.

Olley, J., Caitcheon, G., and Murray, A. (1998). The distribution of apparent dose as determined by optically stimulated luminescence in small aliquots of fluvial quartz: implications for dating young sediments. Quaternary Science Reviews, 17(11):1033– 1040.

Olley, J., De Deckker, P., Roberts, R., Fifield, L., Yoshida, H., and Hancock, G. (2004). Optical dating of deep-sea sediments using single grains of quartz: a comparison with radiocarbon. Sedimentary Geology, 169(3):175–189.

Perilla-Castillo, P. J., Driese, S. G., Horn, S. P., Rittenour, T. M., Nelson, M. S., and McKay, L. D. (2023). Using soil micromorphology to assess the reliability of radiocarbon and osl dating of fluvial deposits. Physical Geography, 0(0):1–53.

Prescott, J. and Hutton, J. (1994). Cosmic ray contributions to dose rates for luminescence and ESR dating: Large depths and long-term time variations. Radiation Measurements, 23(2):497–500.

Roberts, R., Galbraith, R., Olley, J., Yoshida, H., and Laslett, G. (1999). Optical dating of single and multiple grains of quartz from Jinmium rock shelter, northern Australia, part 2, Results and implications. ARCHAEOMETRY, 41(2):365–395.

Singh, A., Thomsen, K. J., Sinha, R., Buylaert, J.-P., Carter, A., Mark, D. F., Mason, P. J., Densmore, A. L., Murray, A. S., Jain, M., et al. (2017). Counter-intuitive influence of Himalayan River morphodynamics on Indus civilisation urban settlements. Nature communications, 8(1):1617.

Technical University of Denmark (2019). Sophia HPC Cluster. Research Computing at DTU.

Thiel, C., Buylaert, J.-P., Murray, A., Terhorst, B., Hofer, I., Tsukamoto, S., and Frechen, M. (2011). Luminescence dating of the Stratzing loess profile (Austria)– testing the potential of an elevated temperature post-IR IRSL protocol. Quaternary International, 234(1-2):23–31.

Thomsen, K., Murray, A., Buylaert, J., Jain, M., Hansen, J., and Aubry, T. (2016). Testing single-grain quartz OSL methods using sediment samples with independent age control from the Bordes-Fitte rockshelter (Roches d'Abilly site, central France). Quaternary Geochronology, 31:77–96.

Thomsen, K. J., Kook, M., Murray, A., Jain, M., and Lapp, T. (2015). Single-grain results from an EMCCD-based imaging system. Radiation Measurements, 81:185– 191.

Thomsen, K. J., Murray, A., and Jain, M. (2012). The dose dependency of the overdispersion of quartz OSL single grain dose distributions. Radiation Measurements, 47(9):732–739.

Thomsen, K., Murray, A., Bøtter-Jensen, L., and Kinahan, J. (2007). Determination of burial dose in incompletely bleached fluvial samples using single grains of quartz. Radiation Measurements, 42(3):370–379.

Thomsen, K. J., Murray, A., and Bøtter-Jensen, L. (2005). Sources of variability in OSL dose measurements using single grains of quartz. Radiation measurements, 39(1):47–61.

Vanderberghe, D., De Corte, F., Buylaert, J.-P., Kučerac, J., and Vanden Haute, P. (2008). On the internal radioactivity in quartz. Radiation Measurements, 43:771– 775.

Wintle, A. G. and Murray, A. S. (2006). A review of quartz optically stimulated luminescence characteristics and their relevance in single-aliquot regeneration dating protocols. Radiation measurements, 41(4):369–391.

Yoshida, H., Roberts, R. G., Olley, J. M., Laslett, G. M., & Galbraith, R. F. (2000). Extending the age range of optical dating using single 'supergrains' of quartz. Radiation Measurements, 32(5-6), 439-446.

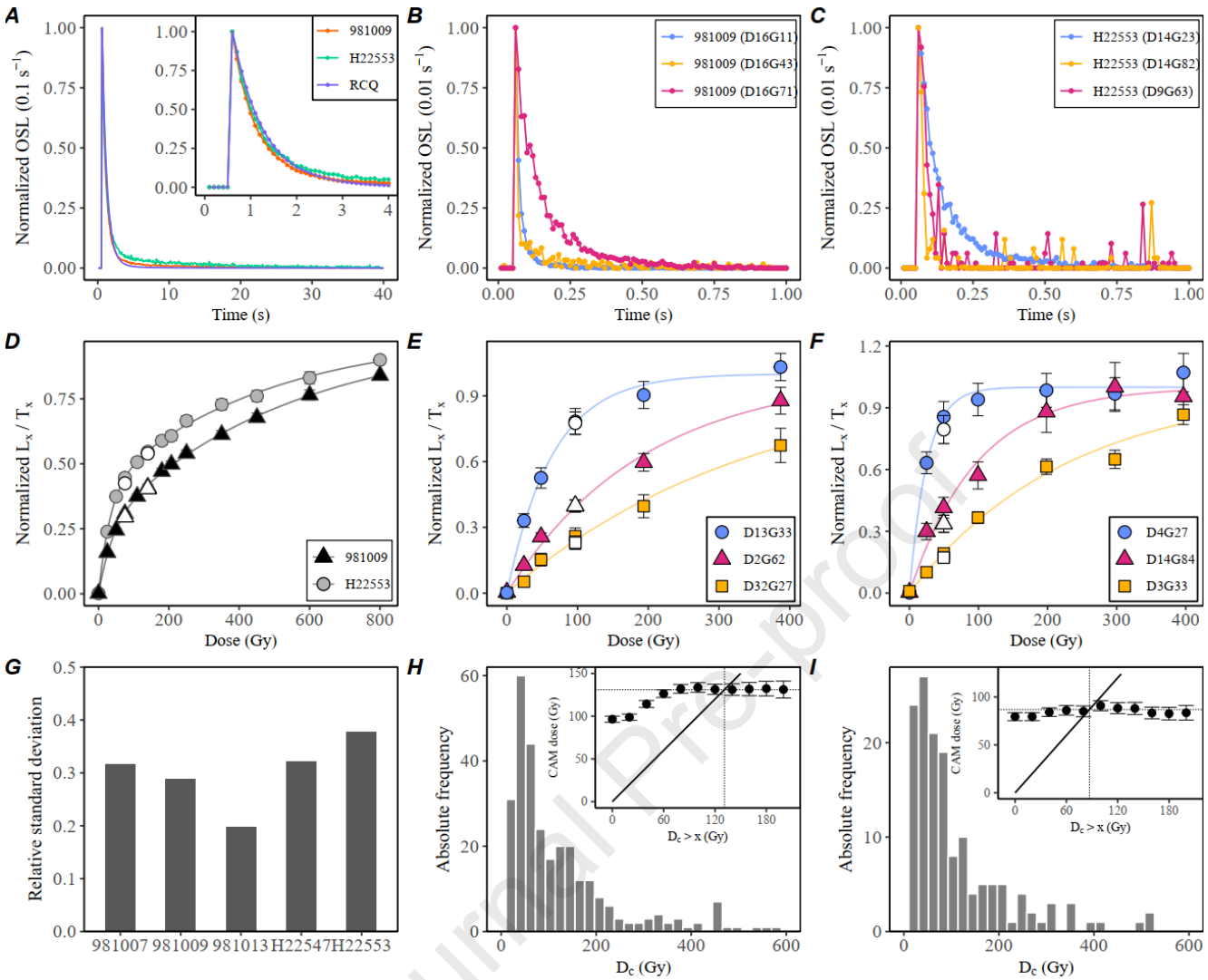


Fig.1. Luminescence characteristics for Gammelmark and Sula samples. Left-hand graphs (A, D, G) show multi-grain (MG) characteristics. Middle and right-hand graphs (B, C, E, F, H, I), show single-grain (SG) characteristics. (A) Representative normalized MG stimulation curves for 981009, H22553 and Risø Calibration Quartz (RCQ). (B) SG stimulation curves for three individual grains from sample 981009. (C) SG stimulation curves for three individual grains from sample H22553. (D) Representative MG dose response curves (DRCs) from Sula (sample H22553) and Gammelmark (sample 981009). Open symbols indicate recycling ratios (E) DRCs of three individual Gammelmark grains (sample 981009). Open symbols indicate recycling ratios (F) DRCs of three individual Sula grains (sample H22553). Open symbols indicate recycling ratios. (G) MG relative standard deviation of natural dose estimates for each sample. (H) SG D_c-distribution from 981009 (n=339). The inset shows the effect of omitting single-grain dose estimates based on their D_c values for CAM doses. (I) SG D_c-distribution from H22553 (n = 104). The inset shows the effect of omitting single-grain dose estimates based on their D_c values for CAM doses.

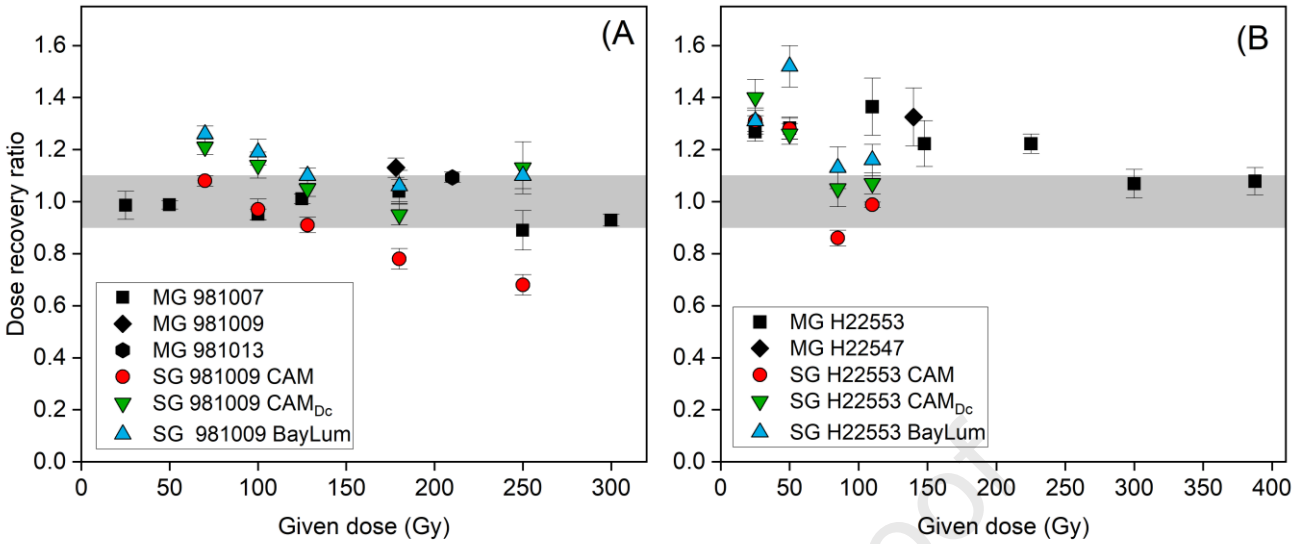


Fig.2. Quartz SAR dose recovery ratios (measured dose divided by given dose) as a function of given dose for A) Gammelmark (981007, 981009 and 981013) and B) Sula (H22547 and H22553). "MG" is multi-grain and "SG" is single grain. Prior to giving any dose, the aliquots were bleached using either blue LED stimulation (MG and SG) or the SOL2 simulator (MG only). No detectable difference is observed between the two bleaching modes and the results are combined (see Fig.SI.3.1 for further information). For "CAM" (red circles) the measured dose has been calculated using the s_{Tn}<20% criterion. For "CAM_{Dc}" (green downward facing triangles) the measured dose has been calculated the using the s_{Tn}<20% criterion in combination with the D_c criterion. For "BayLum" (blue upward facing triangles) the measured dose has been calculated using the s_{Tn}<20% criterion including saturated grains. All SG data is tabulated in Tabel SI.4.

Table 1. Summary of predicted doses and multi-grain quartz natural doses normalized to the age control. "Predicted D_e" is the dose expected from the age control (128±2 ka) and the total dose rate (see Table SI.2). "*n*" is the number of aliquots used in equivalent dose estimation and "n_{sat}" is the number of aliquots which did not give a bounded dose estimate. We show dose estimates for the arithmetic average ("Av."), Median and BayLum. Uncertainties are given at 68% confidence. The brackets for the BayLum results indicate the 68% CI. Individual uncertainties are the propagated uncertainties from the measured and predicted doses.

Sample	Predicted D_e (Gy)	*n*	n_{sat} (%)	Av. D_e	Median D_e	BayLum D_e	
981007	186 ± 8	32	11%	0.92 ± 0.07	0.91 ± 0.08	0.99	[0.94,1.04]
981009	188 ± 9	33	8%	0.98 ± 0.07	0.98 ± 0.08	1.00	[0.89,1.05]
981013	217 ± 10	29	9%	0.84 ± 0.05	0.84 ± 0.05	0.91	[0.86,0.95]
H22547	132 ± 7	26	10%	1.15 ± 0.10	1.07 ± 0.09	1.21	[1.08,1.30]
H22553	95 ± 6	34	3%	1.12 ± 0.10	1.12 ± 0.12	1.21	[1.14,1.27]

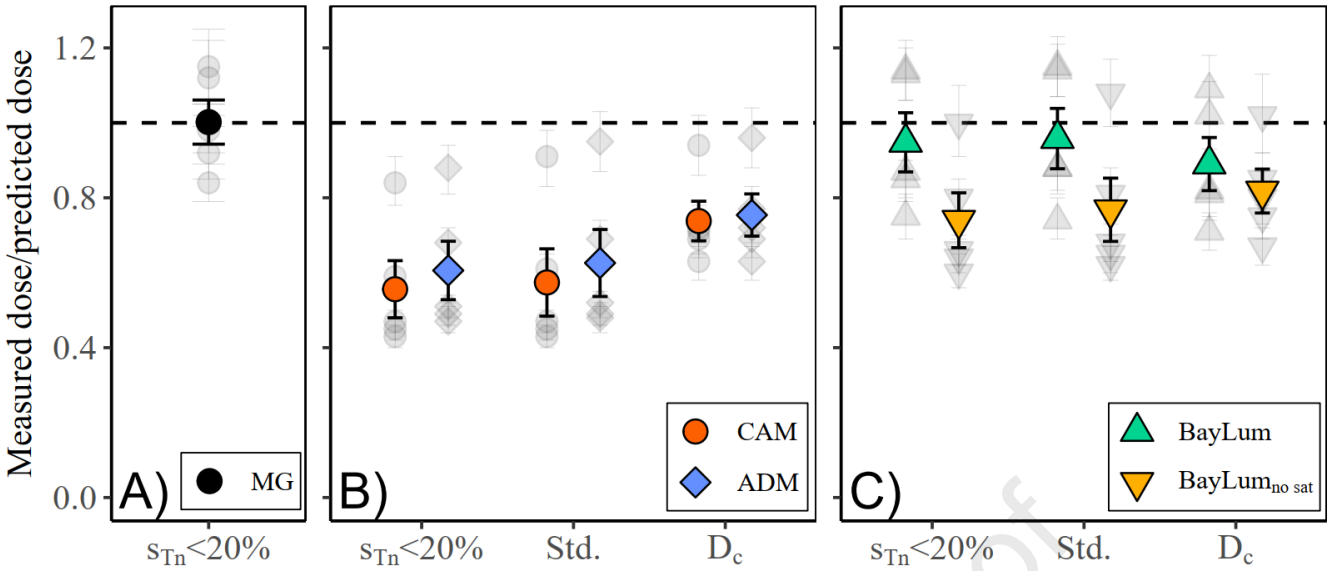


Fig. 3. Ratios of measured equivalent doses to predicted equivalent doses as a function of rejection criteria and dose model. A) Multi-grain quartz OSL results using the arithmetic average. B) Single-grain quartz OSL results using the Central Age Model (CAM) and Average Dose Model (ADM). For ADM, we use an intrinsic OD of 45% for the Gammelmark samples and 30% for the Sula samples (see main text). C) Single-grain quartz OSL results using BayLum - both with the inclusion of grains that do not give bounded dose estimates in frequentist analysis (upward facing triangles) and without these (downward facing triangles). Individual sample ratios are shown as light grey symbols.

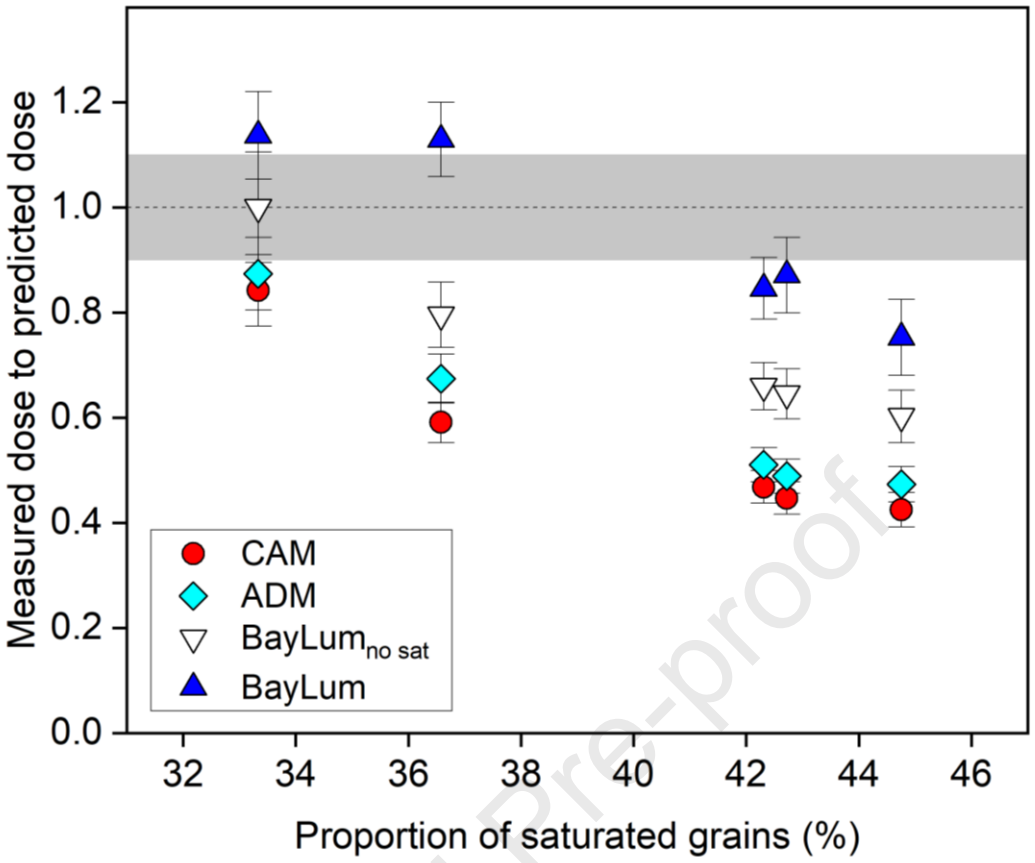


Fig. 4. Ratios of measured to predicted single grain dose as function of the relative number of grains in saturation (%). The data shown are all obtained using the s_{Tn}<20% criterion. The horizontal dashed line indicates unity and the grey band indicates ± 10% of unity.

Table 2. Overview of single-grain results for CAM, ADM and BayLum dose models using different rejection criteria: i) the uncertainty on the natural test dose is less than 20% ("s_{Tn}<20%"), ii) s_{Tn}<20% in combination with the recycling and the IR depletion ratio within 2σ of unity and the recuperation is less than 5% of the natural signal, iii) s_{Tn}<20% in combination with the D_c criterion (see text for details). "*N*" denotes the total number of grains measured while "*n*" represents the number of grains giving bounded dose estimates. "n_{sat}" denotes the number of light-giving grains which did not give bonded dose estimates because of saturation effects. "$BayLum_{no\ sat}$" is the BayLum dose estimated obtained by only including grain giving bounded dose estimates in the analysis, whereas ``BayLum" is the BayLum dose estimate obtained by including grains deemed to be in saturation in the standard frequentist analysis. Results are given at 68% confidence for each of our chosen rejection criteria schemes. Also shown are relative over-dispersion values (OD). For ADM, we use an intrinsic OD of 45% for the Gammelmark samples (9810xx) and 30% for the Sula samples (H225xx).

Sample	*N*	Rejection criteria	*n*	*n/N* (%)	n_{sat} (%)	OD (%)	CAM (Gy)	ADM (Gy)	$BayLum_{no\ sat}$ (Gy)	BayLum (Gy)
H22547	8,900	s_{Tn}<20%	293	3%	37%	58 ± 3	78 ± 3	89 ± 4	105 [99, 109]	149 [146, 154]
		Std.	254	3%	38%	59 ± 4	80 ± 4	91 ± 4	107 [100, 112]	150 [147, 154]
		D_c	158	2%	12%	52 ± 4	93 ± 5	102 ± 5	112 [104, 119]	135 [125, 143]
H22553	6,500	s_{Tn}<20%	104	2%	33%	42 ± 4	80 ± 4	83 ± 4	95 [87, 101]	108 [105, 112]
		Std.	75	1%	39%	41 ± 5	86 ± 5	90 ± 5	102 [97, 109]	109 [107, 113]
		D_c	63	1%	10%	32 ± 4	89 ± 5	91 ± 5	97 [90, 107]	103 [99, 111]
981007	7,000	s_{Tn}<20%	158	2%	45%	64 ± 4	79 ± 5	88 ± 5	112 [104, 118]	140 [128, 147]
		Std.	133	2%	48%	64 ± 5	80 ± 5	89 ± 5	116 [109, 123]	138 [128, 144]
		D_c	64	1%	10%	40 ± 5	117 ± 7	117 ± 7	125 [118, 132]	132 [124, 140]
981009	7,600	s_{Tn}<20%	300	4%	42%	62 ± 3	88 ± 4	96 ± 4	124 [118, 129]	159 [151, 165]
		Std.	252	3%	46%	63 ± 4	88 ± 4	97 ± 4	128 [121, 134]	166 [157, 173]
		D_c	95	1%	8%	46 ± 4	129 ± 7	129 ± 7	141 [132, 150]	154 [142, 163]
981013	4,700	s_{Tn}<20%	240	5%	43%	62 ± 4	97 ± 5	106 ± 5	140 [133, 148]	189 [176, 198]
		Std.	229	5%	43%	60 ± 4	98 ± 5	106 ± 5	141 [133, 149]	191 [178, 201]
		D_c	61	1%	3%	37 ± 5	157 ± 9	157 ± 9	174 [158, 183]	176 [162, 187]

Declaration of interests

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐ The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: