Application of Bayesian analysis to improve the accuracy and precision of optically stimulated luminescence dating techniques

Baumgarten, Frederik Harly

Publication date:
2023

Document Version
Publisher's PDF, also known as Version of record

DTU Physics
RadPhys

Application of Bayesian analysis to improve the accuracy and precision of optically stimulated luminescence dating techniques

A Ph.D Thesis by Frederik Harly Baumgarten

Preface

This dissertation is submitted in partial fulfilment of the requirements for the Ph.D. degree in Physics at the Technical University of Denmark (DTU). The Ph.D. project was initiated on November 1st, 2019, and was completed on December 7th, 2023. The work was carried out at the Luminescence Research Laboratory, Risø, DTU Physics, under the supervision of Senior Scientist Kristina Jørkov Thomsen (main supervisor) and Senior Scientist Mayank Jain (co-supervisor). This project has been partly funded by Danmarks Frie Forskningsfond, Natur og Univers (Independent Reasech Fund Denmark). Grant number 9040-00308B.

The main aim of this work was to test and apply Bayesian analysis to improve the accuracy and precision of optically stimulated luminescence dating techniques and to provide an OSL age chronology for the Upper Palaeolithic site of Kostenki 17, Russia.

Application of Bayesian analysis to improve the accuracy and precision of optically stimulated luminescence dating techniques

Ph.D Thesis
December, 2023

By
Frederik Harly Baumgarten

Cover photo: Midjourney, 2023 (https://www.midjourney.com/app)
Published by: DTU, RadPhys, Frederiksborgsvej 399, Building 201, Risø Campus Denmark
www.fysik.dtu.dk

Acknowledgements

First and foremost, I wish to give special praise, and a separated paragraph, to my girlfriend and life partner, Sara. For all her support, relentlessly "flamey" sense of humor and for being an amazing mother to our wonderful daughter when I could not always be home. Between my five hour daily commutes, external stays and sampling campaigns, I acknowledge the sacrifices she has made for me to attain this degree.

As it goes to the great deal of assistance rendered on the academic side, I absolutely must acknowledge my principal supervisor, Kristina Jørkov Thomsen. It feels obvious, and criminally underwhelming, to state that her footprint in all of this has been instrumental. Kristina as a supervisor was always available, always supportive in times of crisis and I will always remember her labnotes - she is the most meticulous person I have ever met. Guillaume Guérin is another person I wish to thank. I studied abroad at his lab in Rennes, France. As time went on after my stay, Guillaume became more and more involved in all aspects of my project, and his help has been greatly appreciated. I would regret not mentioning that I, as of this writing, have a positive chess score against him. The laboratory technicians at the Risø laboratory, Vicki and Warren, were always helpful and a quintessential source of friendly conversation. Being a Ph.D student, you are subject to an ever changing social environment around you. In my extended stay as a part time student, with parental leave to boot, I have seen most everyone leave Risø a fully fledged Ph.D., and I may be the last of the pre-COVID generation. Thank you to all who was before, and to you, Amélie, who I have had the chance to share office with for a limited time. I also would like to thank Andrew Sean Murray. Here is a guy who seemingly produces two new Ph.D. theses every half hour of conversation. Jan-Pieter Buylaert, who inspired me to one day open a mobile waffle-stand (also he contributed significantly to my study). And also Natalia, who helped me with Russian translations and source-finding.

Your assistance was much appreciated
Frederik Harly Baumgarten

Abstract

Luminescence dating is an important absolute chronological dating technique which determines when a mineral was last exposed to heat or light. It is particular used for sediment burial dating and can extend back ∼500 ka (depending on the environmental dose rate). This makes luminescence dating an incredibly useful tool for a wide array of archaeological, geological and climatological studies. Two major approaches to luminescence dating are quartz single-grain and multi-grain dating. Single-grain dating is sometimes favored over multi-grain dating, even though it has not been tested against independent age control to the same extent as multi-grain dating, particularly not for samples older than 50 ka. Traditionally, luminescence analysis is carried out using a frequentist approach, but recently it has been suggested that Bayesian analysis may be more appropriate. In such an approach all analysis steps are combined into a single model.

The three main objectives of this thesis are 1) to explore the behavior of such a a Bayesian hierarchical model (called "BayLum"), 2) to test the accuracy of single-grain dating using five samples known to have been buried during the early Eemian warm period (∼128 ka), and 3) to provide a robust and absolute luminescence chronology for Kostenki 17, which is an Upper Palaeolithic site potentially recording the first arrival of Anatomically Modern Humans (AMH) in Eastern Europe.

Results show that conventional frequentist single-grain dating analysis underestimates the independent age control by 40-50% on average. However, standard multi-grain dating, and single-grain dating using BayLum, recover the age control within 1σ. I also provide a luminescence chronology for Kostenki 17, and date the first presence of AMH at this site to 48.8 ka, 95% credible interval [46.4 ka, 51.4 ka]. I thus present the earliest evidence of AMH in Eastern Europe.

Resumé

Luminescensdatering er en vigtig geokronologisk dateringsteknik, der bl.a. anvendes til at datere hvornår et sediment blev aflejret. Teknikken kan datere sådanne begivenheder ∼500,000 år tilbage i tiden (afhængigt af hvor meget baggrundsstråling der er tilstede). Dette gør luminescensdatering til et utroligt nyttigt værktøj til en bred vifte af arkæologiske, geologiske og klimatologiske projekter. To store tilgange

til luminescensdatering er: single-grain datering og multi-grain datering af kvarts. Single-grain datering foretrækkes undertiden frem for multi-grain datering, på trods af at metoden er testet mindre mod uafhængig alderskontrol, især med prøver som er ældre end 50 ka. I mellemtiden er en ny udvikling sket inden for luminescence-feltet: fremkomsten af "BayLum" - en bayesiansk hierarkisk model, der kombinerer alle de sædvanlige trin i analysen for at producere en alder fra luminescensmålinger

De vigtigste mål med denne afhandling er at udforske adfærden af BayLum og teste single-grain daterings metoder på fem prøver, der vides at være blevet begravet under den tidlige Eemian varme-periode ($\sim$128 ka). Et vigtigt mål med denne afhandling er at give en luminescens alderskronologi for Kostenki 17, som er en øvre paleolitisk site med potentiel relevans for ankomsten af anatomisk moderne mennesker (AMH) til Østeuropa.

Konventionelle single-grain analysemetoder undervurderede alderskontrol med i gennemsnit 40-50% i dosisområdet 100-200 Gy. Multi-grain og BayLum single-grain kom indenfor 1σ af alderskontrolen. Jeg præsenterer her også en luminescens alderskronologi for Kostenki 17 og daterer tilstedeværelsen af AMH ved dette site til 48.8 ka, 95% credible interval [46,4 ka, 51,4 ka]. Jeg præsenterer dermed de tidligste beviser på tilstedeværelsen af AMH i Østeuropa.

Contents

1 Introduction

1.1 An introduction to Luminescence dating

At the core of Luminescence dating is the existence of a relationship between the amount of energy stored in a mineral, such as quartz or feldspar, and the intensity of the light emitted upon release of that stored energy. The idea to apply this concept to archaeological dating was mentioned for the first time by Daniels et al. (1953), and about a decade later the first application would be published. Aitken et al. (1964) dated an assembly of pottery fragments where they utilized the fact, that when the fragments were first formed, the material had likely been heated to at least 800 ℃ - enough for the luminescence signal to reset. In the time since that original heating event, the pottery would accumulate energy due to continuous exposure to an ever present non-zero level of radiation. They could then release and measure the stored energy in a heating event of their own making. Having measured the light intensity corresponding to the stored energy, the authors then conducted a series of measurements in which they irradiated samples with a known quantity of radiation. After each irradiation, they measured the corresponding light intensity and thus they could construct a dose response curve, and interpolate the light intensity first measured to obtain a measure of the amount of radiation which must have been absorbed in order for the minerals to emit the original light signal measured. Given knowledge about the minerals' yearly absorption of radiation, authors could now calculate an age for each of their materials. And so "Thermoluminescence dating" or "TL dating" was born.

The light that Aitken et al. (1964) were able to measure when they heated their materials came from the release of excited electrons which continuously accumulated in traps/defects of the mineral crystal lattices in the time since the pottery was formed. These electrons became excited through the interaction with radiation absorbed by the minerals over time. They then moved around the minerals where some would became trapped. To empty the traps, a surge of energy was required to

elevate the energy state of the trapped electrons. Aitken et al. (1964) achieved this by supplying heat energy through a heating event. The now freed electrons underwent a de-excitation process, in which the energy state of the electrons was reduced. For some electrons, this meant emitting excess energy as light - the luminescence signal.

To induce a luminescence signal, the energy trigger does not have to be heat. Some electron traps are sensitive to light - a fact which would later enable us to date when sediment was last exposed to sunlight. Sunlight will reset the luminescence signal of minerals on the surface, but when buried, the minerals will continuously build up a luminescence signal. Huntley et al. (1985) introduced a method which exploited light-sensitive traps to date when sediments were buried. Using an argon-ion laser, these authors were able to optically stimulate sediment quartz from a variety of sites and measure the resulting luminescence signal. Their results showed good agreement with radiocarbon and TL age controls.

Another milestone was reached when Murray and Wintle (2000) introduced the "single-aliquot regenerative-dose protocol" (SAR protocol), in which each measurement of a luminescence signal is sensitivity-corrected by following up this first measurement with a secondary irradiation of a fixed dose - a test dose - and then a luminescence measurement of said test dose. Their work aimed to solve the issues that came with preheating minerals before measuring the luminescence signal, as it had been shown that preheating caused sensitivity changes in quartz (Wintle and Murray, 1998, 1999). It is important to mention that the preheating step itself is necessary to empty electron traps that are not stable over geological time scales, as mentioned in the original work of Huntley et al. (1985), who preheated each of their quartz extracts to 250 ℃. The SAR protocol would become standard in OSL dating.

TL and OSL dating methodologies have refined and diversified since the years of their invention, and for a more comprehensive story of the early development of Luminescence dating, I refer to Wintle (2008).

1.2 Constructing a luminescence chronology

An advantage of luminescence dating is that the material which is to be dated is abundant and occurs most everywhere on land and throughout the sediment profile. Consequently, a luminescence researcher is able to construct a chronology for the

entire sediment column. But to do this, decisions must be made along the way.

1.2.1 Multi-grain or single-grain?

Original OSL applications measured the cumulative signal of a collection of grains (multi-grain aliquots, Huntley et al., 1985; Godfrey-Smith et al., 1988; Rhodes, 1988). But later developments would make single-grain measurements not only possible - but also feasible (Duller et al., 1999; Bøtter-Jensen et al., 2003a). The usefulness of single-grain dating is showcased in the case study of the Jinmium rock shelter in Australia. Here, original TL dating suggested human occupation before 116 ± 12 ka (Fullagar et al., 1996), while subsequent single-grain OSL dating efforts revised this age to around 10 ka (Roberts et al., 1998; Galbraith et al., 1999; Roberts et al., 1999). The latter papers became heralded as being among the most important contributions to archaeometry (Wintle, 2008).

When sample materials have been bleached only partially (i.e., incomplete bleaching), single-grain measurement of individual grains may help diagnose this issue and by using statistical tools it is possible to identify the grain population that was fully bleached at deposition (e.g., Duller, 1994, 1995; Olley et al., 1998). Multi-grain measurements, on the other hand, will record a cumulative signal of all the grains, which can lead to an overestimation of the burial dose if incomplete bleaching is present (e.g., Li, 1994; Rhodes and Pownall, 1994; Murray et al., 1995; Berger, 1995; Olley et al., 1998; Duller, 2008). Murray et al. (2012) investigated 67 young quartz samples from a variety of both colluvial and fluvial environments, an found that samples contained an average residual dose of 2 Gy. This would indicate that, on average, only young samples (<20 ka) will be significantly affected by incomplete bleaching. Despite this, single-grain dating is routinely used to date much older samples (e.g., Jacobs et al., 2008; Armitage et al., 2011; Arnold and Roberts, 2011; Jacobs et al., 2011; Arnold et al., 2013; Jacobs and Roberts, 2015; Demuro et al., 2019; Liu et al., 2022).

Single-grain dating can be preferred over multi-grain dating for reasons that do not concern incomplete bleaching. For another attractive feature of single-grain analysis, is the ability to evaluate the OSL characteristiscs of each individual grain. In the search for the "*true*" paleodose, it seems advantageous to be able to identify grains that show undesirable OSL characteristics and omit them from dose calculations. For example, using a SAR protocol, can we really trust the dose-interpolation of a

grain for which repeated (same regenerative dose) test-dose corrected signals give significantly different results? Over the years, several different criteria (rejection criteria) have been used to weed out these potentially "bad" grains. Objective criteria for keeping a grain in the analysis commonly take the following general forms:

- a grain must emit a detectable test dose signal which enables the measurement of meaningful dose response curves. Examples are grains whose relative test dose signal uncertainty is less than 20% ($s_{T_n} < 20\%$), or grains whose test dose signal is at least three times that of the test dose background signal.
- the ratio of test-dose corrected signals from repeated regenerative steps of the SAR (recycling ratio) should be close to one. One example: unity must be within 2σ of the recycling ratio.
- the ratio of test-dose corrected signals from two regenerative steps given the same regenerative dose, with an extra IR stimulation step before the latter OSL stimulation (IR depletion ratio) should be close to one. Unity must be within 2σ of the IR depletion ratio.
- a 0-dose regenerative dose step of the SAR must produce a small test-dose corrected signal. One example: 0-dose test-dose corrected signal must be less than 5% of the natural test-dose corrected signal.

I will here introduce yet another criterion besides the ones listed above: the D_c criterion (Thomsen et al., 2016). The D_c criterion seeks to reject grains whose dose response curve characteristics would not allow them to record the dose accumulated since burial. This criterion focuses on the curvature parameter, D_c, of the following saturating exponential function:

$$L_x/T_x = A \times (1 - exp(-D/D_c))$$

The higher the D_c, the higher the dose needed for the above function to reach the saturation defined by the A parameter. Of course, the burial dose is not known beforehand, and so one cannot, at the outset, specify a sensible lower value limit of D_c that all grains must minimally possess. The way in which this D_c threshold is determined in all relevant chapters of this thesis is by the following:

For each sample:

1. Calculate the apparent dose for grains filtered by an increasing D_c-threshold.
2. Plot the apparent dose as a function of D_c-threshold. Generally, the apparent dose calculated from all grains will be of higher magnitude initially than the D_c-threshold of 0 Gy. Where the D_c-criterion will be especially useful, is when the threshold is increased and as a result, the apparent dose also increases.
3. Calculate the intersect of the apparent dose on the 1-to-1 line between the apparent dose and the D_c-threshold.
4. All grains must have at least the value determined in (3) as the D_c-value.

Figure 1.1 illustrates this process. Figure 1.1A shows a sample with a given dose of 50 Gy (dose recovery experiment). In this case, the application of a D_c-threshold will do little else than reduce the number of grains used to estimate the central dose. The central dose itself is unaffected. In Figure 1.1B another sample was given a dose of 250 Gy (dose recovery experiment). In contrast with the previous dose-recovery experiment, application of the D_c-threshold now greatly affects the estimated dose. The idea is, when no bias towards lower doses exists in the low-D_c grains, the D_c criterion will not change central estimates. If such a bias does exists among the lower-D_c grains, the D_c criterion will change central estimates. The D_c criterion is particularly useful in samples where there are a lot of grains where the natural sensitivity corrected signal cannot be interpolated onto the laboratory dose response curve because of saturation effects. The worry is not so much these particular grains, which cannot be included in a standard frequentist dose analysis, but rather the grains which is on the low tail of the light distribution that will just manage to interpolate onto the dose response curve nonetheless and hence produce a too small estimate of the central dose which they are not able to record.

Multi-grain dating of quartz does have one advantage over single-grain dating of quartz: history. Over the years, multi-grain dating has been tested against independent age control and been shown to be accurate over a wide range of ages (e.g., Murray et al., 2021). For single-grain dating of quartz, especially with samples older than 50 ka, these types of studies are scarce.

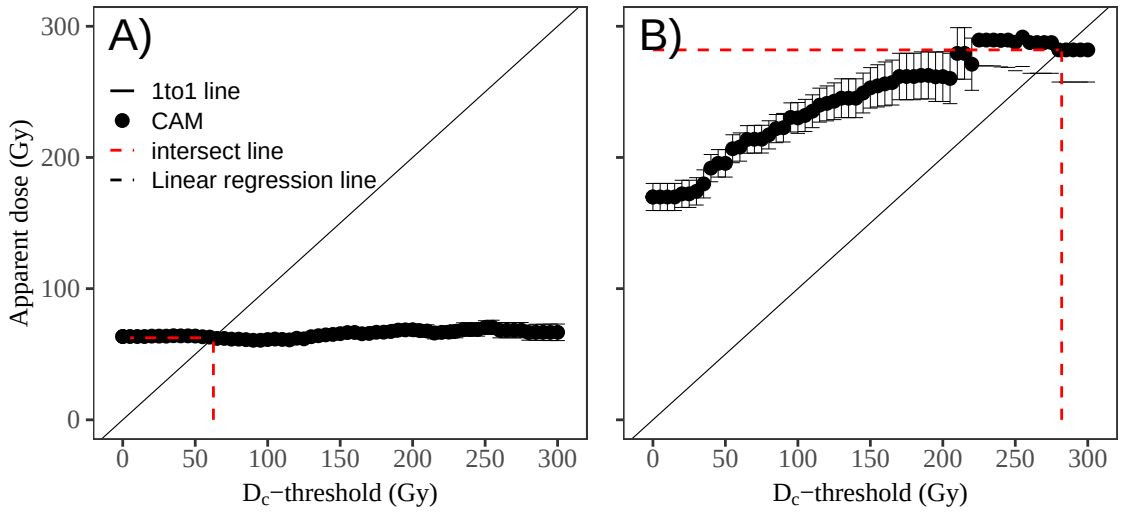

Figure 1.1: Apparent dose (Gy) as a function of D_c-threshold. (A): Dose recovery with a given dose of 50 Gy (sample code H22553). (B): Dose recovery with a given dose of 250 Gy (sample code 981009). Apparent dose is in this case calculated using the Central Age Model (CAM, Galbraith et al., 1999)

1.2.2 Bayesian invasion: BayLum and Age-depth modelling

To build a chronology, the researcher must choose a framework for converting measured luminescence signals for each measured unit of a sample into a central dose estimate and then an age given information about the dose rate experienced by the sample. This has conventionally been achieved using a multi-step process, with each step of analysis disjointed from the next, culminating in a frequentist estimate of dose. These steps involve constructing a dose response curve from luminescence signals, interpolating the natural signal onto the dose axis, estimation of a central dose and finally, division with the dose rate to obtain an age. A recent development within the field of luminescence is the emergence of "BayLum", which is a dose and age model that combines all the previously mentioned steps into a single Bayesian hierarchical model ("BayLum", Combès et al., 2015; Combès and Philippe, 2017). The way BayLum works is through Bayes Theorem (Equation 1.1), with priors specified for dose response curves and dose model scale and shape parameters along with a few others (Combès et al., 2015). A likelihood function is then defined and a Markov Chain Monte Carlo (MCMC) procedure is used to obtain samples from the posterior. I go into more detail about fundamental Bayesian concepts and MCMC in the following sections. BayLum shows promising initial results (e.g., Guérin et al., 2015; Heydari and Guérin, 2018; Lahaye et al., 2019; Carter et al., 2019; Heydari et al., 2020; Guérin et al., 2020; Heydari et al., 2021; Guérin et al., 2022, 2023).

If the objective is to produce an OSL age at a depth of 402 cm. Which is to prefer: to obtain exactly one age from a depth of 402 cm? Or to obtain a series of ages at around 402 cm? Besides painting a better picture of sample variance, the latter option allows building a relationship between age and depth. With age-depth modelling, it is possible to predict ages for depths not sampled, but it is also likely to increase the precision of each prediction given enough data to inform the modelling procedure. The realm of age-depth modelling has become dominated with Bayesian procedures (Bayliss, 2015), and perhaps with good reason (Trachsel and Telford, 2017). At the forefront of these model procedures are softwares such as OxCal (Ramsey, 2008; Bronk Ramsey, 2009) and Bacon (Blaauw and Christen, 2011a) - both of which model sediment accumulation and employ MCMC routines to produce posterior samples to allow inferences of age at all depths in question. Another interesting model to mention is the ArchaeoPhases age-depth model (Philippe and Vibet, 2020). The ArchaeoPhases age-depth model takes a different approach, producing age predictions by applying local polynomial regression to the relationship between depth and the age parameters of each link of an MCMC chain - the output of Bayesian procedures such as the BayLum age model, or age-depth models such as OxCal and Bacon.

1.3 What are Bayesian statistics?

A sunbird enthusiast gets into a heated argument about the adult female beak length of a particular species of sunbird. Motivated to show that the average beak length is no more than 23.2 mm, the sunbird enthusiast goes to the local museum. Here, our enthusiast is presented with 8 adult female specimens and proceeds to measure each of the bird's culmen (upper beak ridge) and produces the observations of Table 1.1.

Table 1.1: Culmen lengths of 8 adult female mystery sunbirds

Culmen (mm)
24.4
24.9
21.2
21.6
19.4
20.6
24.3
20.4

The sunbird enthusiast now calculates a mean of 22.1 mm along with a measure of uncertainty - a 95% confidence interval: 95% CI [20.3 mm, 23.9 mm]. Within this frequentist approach, the interpretation becomes that if the experiment was completed 100 times, then the true population mean should be contained within the computed 95% confidence interval 95 times out of 100. However, the probability that the one interval which was actually sampled contains the true mean is either 0% or 100%, and there is no way to determine which. Our enthusiast is instead 95% confident that the true mean is within 20.3 mm and 23.9 mm - much to the dismay of our enthusiast, the interval includes values above 23.2 mm. If only there was a way to include more information. Our enthusiast goes for another statistical paradigm: Bayesian statistics. At the heart of the Bayesian paradigm is "Bayes Theorem" (1.1).

$$Posterior = \frac{Likelihood \times Prior}{Normalization} \tag{1.1}$$

In the case of the beak length of the mystery sunbird, the sunbird enthusiast would be determining the probability distribution ("*Posterior*") for the *mean* beak length parameter. This involves calculating the product of the probabilities of observing a particular mean value given each observation ("*Likelihood*") times the "*Prior*" probability of the mean to take on a particular value given the chosen prior. Our enthusiast chooses a normal probability distribution for the likelihood, assuming a known variance of 4 mm^2 (information is added, reasonable or not). For the prior, our enthusiast remembers the correspondence with an author of a certain excellent

master thesis from a few years ago (Baumgarten, 2019), in which the female beak length of a very similar-sized sunbird could be described as normal with mean = 22.5 mm, and variance = 5.5 mm^2 (information is added). When both the likelihood and the prior are normal distributions, the posterior distribution will also be a normal distribution, described by the mean μ_p and the variance σ_p^2. This is a situation which can be solved analytically. First a summary of known quantities:

- Number of sampled observations: n = 8
- Mean of sampled data: $\mu_{meas} = 22.1$ mm
- Mean of prior: $\mu_{prior} = 22.5$ mm
- Variance of prior: $\sigma^2_{prior} = 5.5$ mm^2
- Known variance: $\sigma^2 = 4.0$ mm^2

The posterior is then:

$$P(x) = \frac{1}{\sqrt{2\sigma_p^2\pi}} \times e^{-\frac{(x-\mu_p)^2}{2\sigma_p^2}} \tag{1.2}$$

where the parameters are given as:

$$\mu_p = \frac{\sigma^2\mu_{prior} + n\sigma^2_{prior}y}{\sigma^2 + n\sigma^2_{prior}} = 22.1\,\text{mm}^2 \tag{1.3}$$

and

$$\sigma_p^2 = \frac{\sigma^2\sigma^2_{prior}}{\sigma^2 + n\sigma^2_{prior}} = 0.5\,\text{mm}^2 \tag{1.4}$$

To conclude this Bayesian example, our sunbird enthusiast derives a mean length of 22.1 mm, 95% credible interval [21.2 mm, 23.0 mm] - which argues in favor of the sunbird enthusiast's claim. This time the appropriate interpretation is, that with 95% probability (given the data), the mean is within 21.2 mm and 23.0 mm.

The Bayesian situation above is deliberately a simple one. Real-world situations are often far more complex. For example, imagine that it is not the mean beak length that we desire to make inferences about, but instead the central estimate of OSL doses (given some dose distribution), where the variance (or scale parameter) of the doses cannot be said to be known. This will require another prior term on the dose variance. Perhaps it is also advantageous to include the interpolation of individual

doses from their respective dose response curves into the model? This will add priors on the parameters of the DRCs. Perhaps the likelihood and the priors are not all best defined as normally distributed? We now have a situation where it is not so easy to derive a posterior density distribution. Part of the equation will involve integrating the product of the *Likelihood* and the *Prior* over all possible values of all parameters (The *Normalization* term in Bayes Theorem) - and this can quickly become an intractable task. But there is a way to approximate the solution - using Markov Chain Monte Carlo (MCMC).

1.4 Basics of MCMC

Markov Chain Monte Carlo (MCMC) is one possible way of approximating the multivariate posterior probability distribution, and it does so by continuously generating samples until eventually, the density distribution of the drawn samples mirrors the true posterior. The very first step in MCMC will be to generate an initial value for all parameters of the model. Given this set of values, an MCMC sampler will draw candidate/proposed values which the sampler will either accept or reject. Accepted values will then form the new starting point for the next MCMC iteration. The details of how this is done will depend entirely on the nature of the MCMC sampler that is used. One example of an MCMC sampler is the Metropolis-Hastings algorithm, which generates candidate values from distributions parameterized using the current set of MCMC values. The un-normalized posterior densities is then calculated for both the current and candidate parameter values and the ratio of the two will have an impact on the probability that the new set of values are accepted or rejected. If accepted, the new set of values will influence the draw of the next set of candidate values.

To illustrate the point of the MCMC, imagine a simple Bayesian model containing only two parameters, both of which we want to make inferences about:

- "D": the location parameter of our chosen equivalent dose function.
- "sD": the scale parameter of our chosen equivalent dose function.

The multivariate posterior is then 2-dimensional, conveniently allowing me to plot the hypothetical true posterior density distribution in 2D (Figure 1.2). In this figure, we see regions of varying density, with a high-density hub around D $\approx$ 80 and sD $\approx$

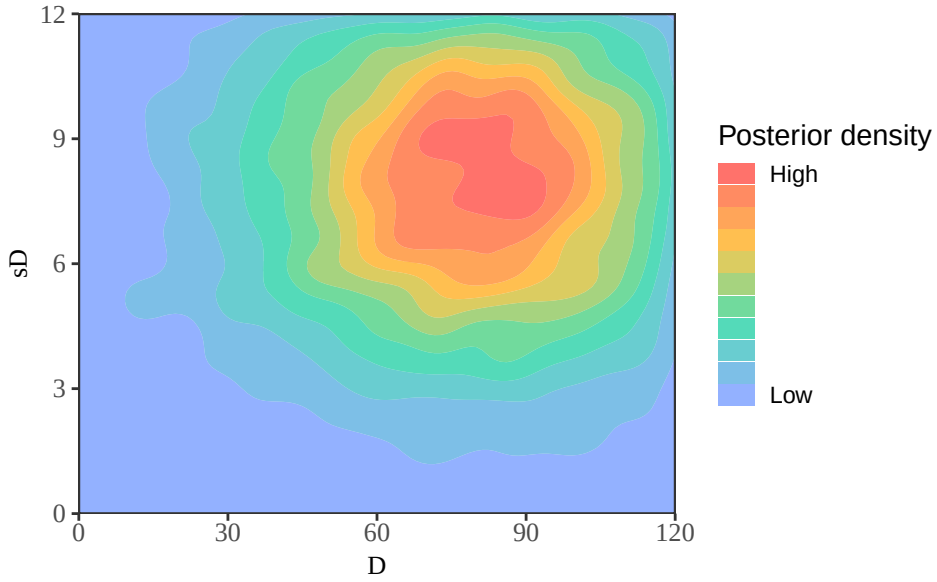

Figure 1.2: Illustration of the true 2D posterior density distribution in an imagined Bayesian model example.

8.

In this imagined example, the initial starting point of the MCMC happens to be in a low-density area (Figure 1.3A). The MCMC sampler will now iteratively generate new samples and move to that parameter coordinate if accepted (Figure 1.3B).

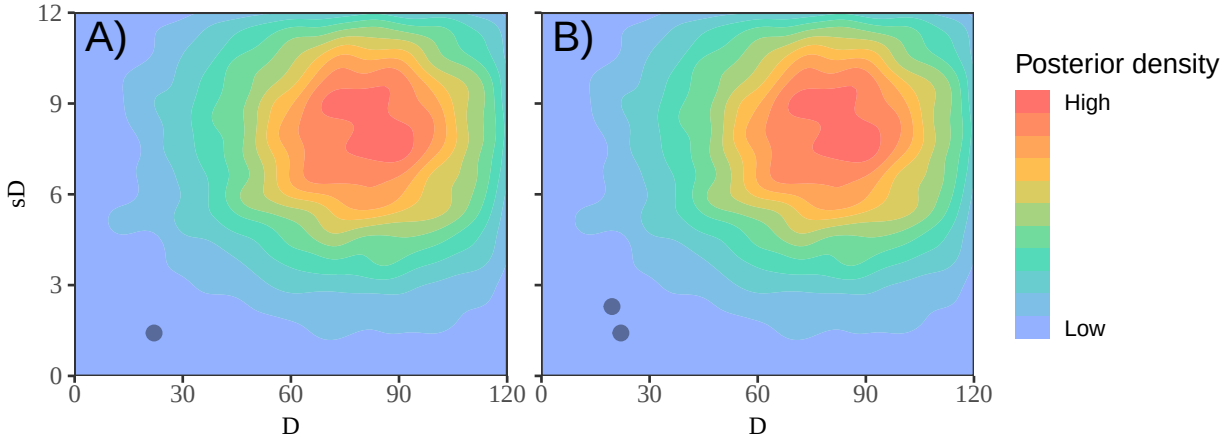

Figure 1.3: Illustration of MCMC search in a 2-parameter Bayesian model. (A) The initial combination of values. (B) The first accepted new values for "D" and "sD"

The nature of the sampling will predispose the sampler towards accepting more probable values (Figure 1.4A and B).

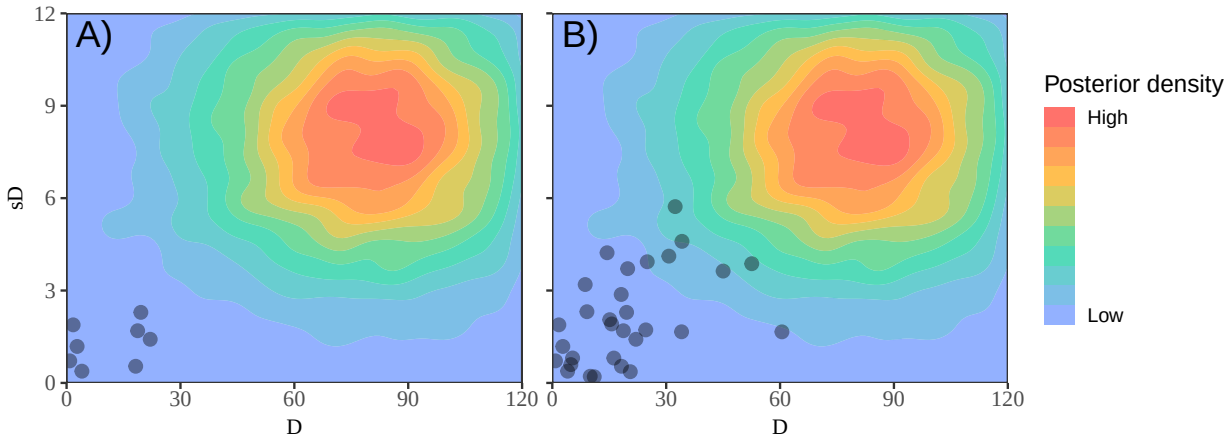

Figure 1.4: Illustration of MCMC search in a 2-parameter Bayesian model. (A) More candidate values are proposed and accepted. (B) Sampler is prediposed into accepting more probable values.

Hopefully, our MCMC will reach the high posterior density region. When this happens, the distribution (shape and location) of accepted candidate values for each parameter will begin to not change by continued sampling of the MCMC sampler, i.e., the location and shape parameters of the marginal posterior distributions (each individual parameter distribution) remain fixed. The distributions become stationary.

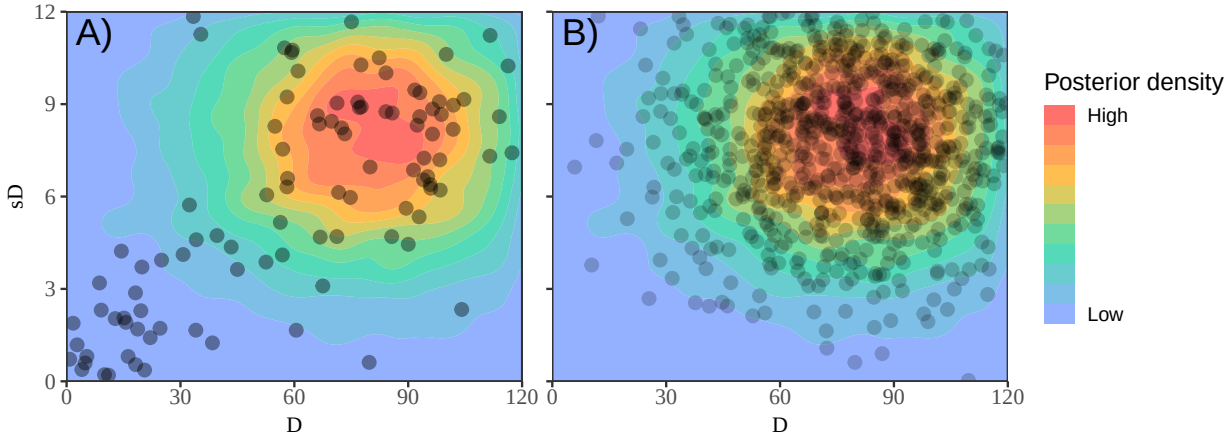

Figure 1.5: Illustration of MCMC search in a 2-parameter Bayesian model. (A) The sampler has started sampling the high-density region of the posterior. (B): Sampler starts sampling the posterior target distribution. We here regard the initial samples as "burn-in", and remove them for our analysis of the posterior.

Figure 1.5A shows that sampling has begun in the high-density region, and in Figure 1.5B, many additional samples have been drawn from this very region. Commonly, the initial set of samples are discarded as "burn-in" Figure 1.5B.

When the marginal posteriors reach a stationary point, we can draw summary conclusions about the parameters of interest given the samples from this stationary distribution (i.e., the mean, median and credible intervals). But how do we know that this point has been reached?

Samples from a single MCMC sampler trying to reach the target distribution constitutes a single chain, with each link representing a combination of values for all model parameters. But if our MCMC runs several individual samplers for the same problem, we can compare each generated chain with each other. There is randomness intrinsic to the MCMC sampling process, and so each MCMC sampler is likely to take a unique path towards the target stationary distribution. It then follows that if recent iterations of samples by each chain produce the same distribution, then the target stationary distribution has been reached. This assessment can be accomplished through many means. In this thesis, I base this evaluation on visual assessment of the marginal posterior distributions and the Rubin-Gelman diagnostic (Gelman and Rubin, 1992).

1.5 Aims and objectives

Luminescence dating is an important geochronological dating technique used to date the burial of sediment going back ∼500 ka (depending on the environmental dose rate, Murray et al., 2021). This makes luminescence dating an incredibly useful tool for a wide array of archaeological, geological and climatological projects. Multi-grain (MG) and single-grain (SG) OSL dating of quartz are two major methodologies within luminescence dating. But while MG OSL dating of quartz has been compared with independent age control for samples originating anywhere between the Holocene and the upper-Pleistocene, little is known about the accuracy of single-grain dating when the burial age extends beyond 50 ka. A recent development in the field of luminescence is the emergence of "BayLum" - an age and equivalent dose model based in a Bayesian hierarchical model framework. Between BayLum and available Bayesian age-depth-modelling software, the of aim of this thesis is to provide an OSL age chronology for the Upper Palaeolithic site of Kostenki 17, and to test and apply Bayesian methods to improve the accuracy and precision of luminescence dating techniques. My objectives are:

1. To explore the behavior of BayLum and improve upon the practical aspects of running BayLum with single samples and high resolution dating in mind.

2. To test and compare multi-grain and single-grain OSL dating of quartz on samples with known doses in the 100-200 Gy dose range, while exploring the use and accuracy of BayLum and comparing with conventional frequentist single-grain methodology.

3. To apply Bayesian methods in order to construct an OSL dating chronology for a very important site presumed to document the first appearance of Anatomically Modern Humans (AMH) on the East-European Plain.

Findings within this thesis will impact future use of BayLum and reveal significant implications of using single-grain OSL dating of quartz in the 100-200 Gy dose range. We will also provide an age chronology for Konstenki 17, Russia, which significantly changes what we know about the arrival of AMH in East Europe.

1.6 Thesis outline

Below I outline each of the five chapters that follow this one:

Chapter 2. Insights from BayLum

In this chapter, I present the results of experiments I performed on the BayLum dose model. I undertook these experiments due to questions I developed while using BayLum to derive OSL ages for a large set of OSL samples. Particularly, I focus on practical aspects of running BayLum (especially when many OSL samples are involved), "general use" and on behavioral aspects of BayLum given extreme data sets.

Chapter 3. Reducing computation time in the R-package 'BayLum'

Here I present recent features implemented into the BayLum R-package. These features primarily seek to improve the required runtime of BayLum. This work has been published in *Ancient TL* (Baumgarten et al., 2023).

Chapter 4. Testing the accuracy of single-grain OSL dating on Eemian quartz samples

Chapter 4 provides a test of accuracy for single-grain OSL dating of Eemian age samples, with expected equivalent doses in the 100 - 200 Gy dose range. Results of conventional single-grain dating methodology are compared with both multi-grain OSL dating of quartz and newer single-grain methodology. This work was presented as an oral presentation in the LED2023 conference (June, 2023) and has been submitted for publication in a reduced form to the LED2023 special edition of *Quaternary Geochronology* (Baumgarten et al., submitted).

Chapter 5. Establishing an OSL chronology for Kostenki 17 - an Upper Paleolithic site by the Don River, Russia

In this chapter, I derive an absolute chronology for Konstenki 17, which is an Upper Palaeolithic site located by the Don River, Russia. Using BayLum and Bayesian age-depth modelling, I present an OSL age chronology for 40 OSL samples obtained between 2017 and 2021.

This chapter will be submitted for publication in an international peer-reviewed journal as Baumgarten et al.

Chapter 6. Summary, conclusions and outlook

In this final chapter of my thesis, I will provide my thesis summary and conclusions. I will also explore future avenues of reseearch.

2 Insight from BayLum

Chapter outline

In this chapter, I present the results of experiments I performed on the BayLum dose model. I undertook these experiments due to questions I developed while using BayLum to derive OSL ages for a large set of OSL samples. Particularly, I focus on practical aspects of running BayLum (especially when many OSL samples are involved), "general use" and on behavioral aspects of BayLum given extreme data sets.

2.1 Introduction

"BayLum" is a dose and age model which takes a Bayesian approach to central dose estimation (Combès et al., 2015; Combès and Philippe, 2017) and is currently made freely available as an R-package (Philippe et al., 2019; Christophe et al., 2023) for use within the statistical coding environment "R" (R Core Team, 2023). A conventional workflow to produce an age from measured OSL signals is the following:

1. Define summation intervals and calculate L_n/T_n and L_x/T_x for each aliquot or grain
2. Construct a dose response curve relationship, relating an OSL signal response (L_x/T_x) with regenerative doses
3. Calculate the dose at which the L_n/T_n-signal intersects the dose response curve for each individual aliquot/grain
4. Calculate a central dose estimate based on a collection of doses obtained in (3)
5. Obtain an age by dividing the central dose estimate from (4) with the environmental dose rate of the sample in question

In BayLum, steps (2) - (5) are not broken up, but instead modelled together within a Bayesian hierarchical framework. The effect of this is to reduce the risk of losing

information, which can be incurred at each step of the above analysis (Combès et al., 2015). For example, consider a dose interpolated on the curving part of the dose response curve using concentional methods: is the symmetrical value of uncertainty given to this dose (here I mean to say that the uncertainty extends an equal distance above and below the estimate) likely to be truly symmetrical as it is input into, for example the CAM dose model?

The major downsides of BayLum probably lies in the practicality of using it: (i) "R" is currently required to run BayLum, which presents a learning curve of its own. (ii) target users may not be familiar with Bayesian hiearchical modelling and coupled with the underlying MCMC processes, BayLum can feel somewhat like a "black box" (iii) BayLum requires a lot of computation time.

In this chapter, I primarily answer questions I developed while applying BayLum to a large set of OSL samples. My hope is that I am able to alleviate some of the aforementioned downsides for future users of BayLum.

2.2 Data set introductions

In the following sections, I explore the behavior of BayLum using both real and simulated OSL data sets. I will now introduce these OSL data sets, which contain a multi-grain sample and two single-grain samples. I also simulate a data set using Risø Calibration Quartz (RCQ).

2.2.1 Real data sets

Multi-grain OSL data sets most often contain far less individually measured units (aliquots) compared with single-grain analysis. The multi-grain data set I have chosen (207733) is from a Russian archaeological site referred to as Konstenki 17 (K17, see Chapter 5). It is an OSL data set consisting of individual dose measurements on 23 aliquots. The obtained dose distribution is approaching a normal distribution with individual dose estimates ranging between 32 and 57 Gy (Figure 2.1).

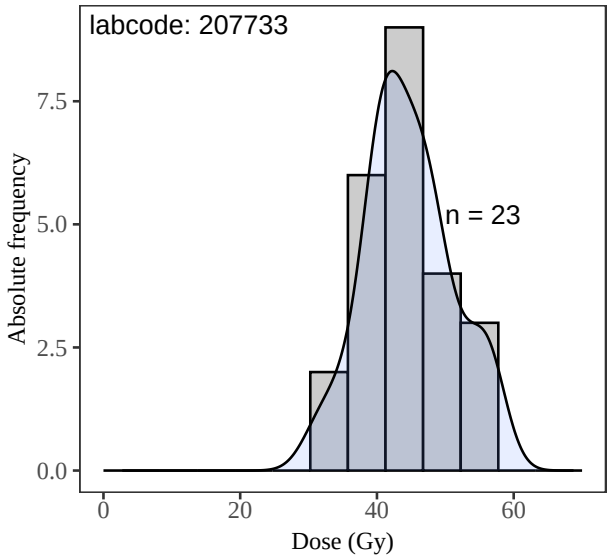

Figure 2.1: Histogram and density plot of a measured multi-grain quartz dose distribution (labcode: 207733).

I test two single-grain data sets: one from K17, Russia (207749, Figure 2.2A), and one from Gammelmark, Denmark (981009, Figure 2.2B). Both data sets contain 300 or more grains which give bounded dose estimates with frequentist approaches and both show dose distributions characterized by a positive skew. However, the two samples are visibly different in terms of their degrees of skewness, and while 207749 has an over-dispersion (OD) of about 40%, 981009 has an OD of about 60%.

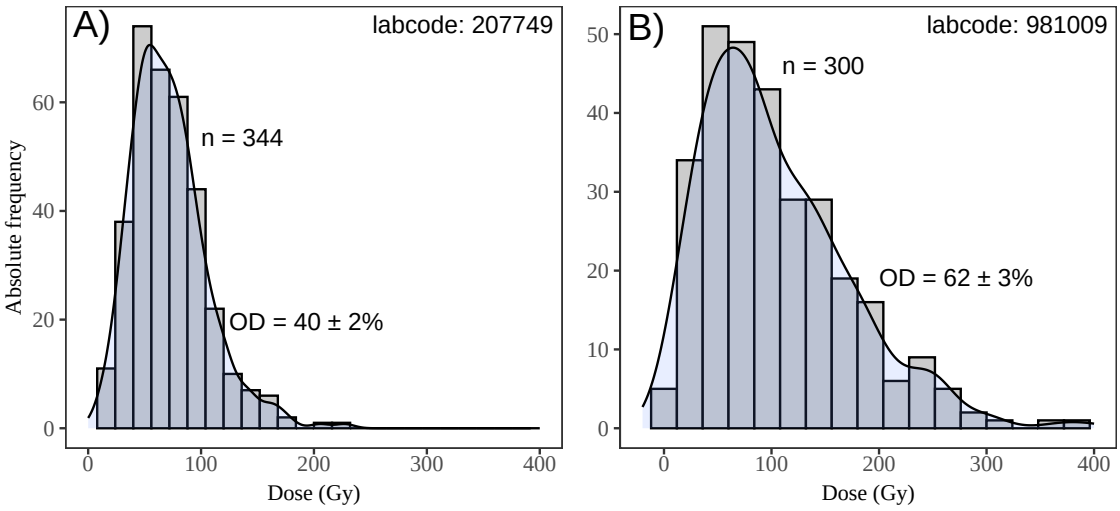

Figure 2.2: Histogram and density plot of two single-grain (SG) quartz dose distributions. (A): Labcode: 207749. A 344 grain SG sample with an over-dispersion (OD) of $40 \pm 2\%$ (B): Labcode: 981009. A 300 grain SG sample with an OD of $62 \pm 3\%$.

2.2.2 Simulated data set

In order to test BayLum with a simulated sample with known dose and over-dispersion, it is not enough to simply generate D_e values together with a plausible set of uncertainties, as it may be in frequentist analysis using e.g., the Central Age Model (CAM, Galbraith et al., 1999) or the Average Dose Model (ADM, Guérin et al., 2017). This is because the BayLum model does not begin with the analysis of individual dose estimates to determine a central dose of all grains. Instead, the BayLum model begins with the information required to build dose response curves for each aliquot or grain. The main data-object containing this information (which is a required input of all the different BayLum modelling functions) is the "DataFile"-object (Christophe et al., 2023). The simplest way to build this object is to use the "Genereate_DataFile()" function within the BayLum-package, and this will require the information to be stored in "bin" or "binx"-files. There is then at least four non-excluding modes of simulating an OSL data set: (i) by using real measurement files ("bin" or "binx"-files) and building grain distributions through selection of the actually measured grains to fit desired characteristics. At least one such example can be found in the literature (Heydari et al., 2020). (ii) using "R" to modify an existing bin-file such that, for example, the natural test dose measurement of grain 1 becomes the natural test dose measurement of grain 3. (iii) by generation of a synthetic binx file using available tools such as R-packages "Luminescence" (Kreutzer et al., 2022), "RLumModel" (Friedrich et al., 2022) and "sandbox" (Dietze and Kreutzer, 2022). (iv) by manipulation of the "DataFile"-object itself.

For these experiments I produced a single simulated sample with an OD of ∼20% by randomly generating observations from a normal distribution ("Template", Figure 2.3B). I then matched each observation with the grain to which the Euclidean distance was smallest from a large pool of bleached Risø Calibration quartz grains given a wide range of doses. These grains (and all their binx-file records) was then added to a new binx file, built sequentially from all the matched grains. As seen from Figure 2.3B, the generated template and the constructed data set is virtually identical.

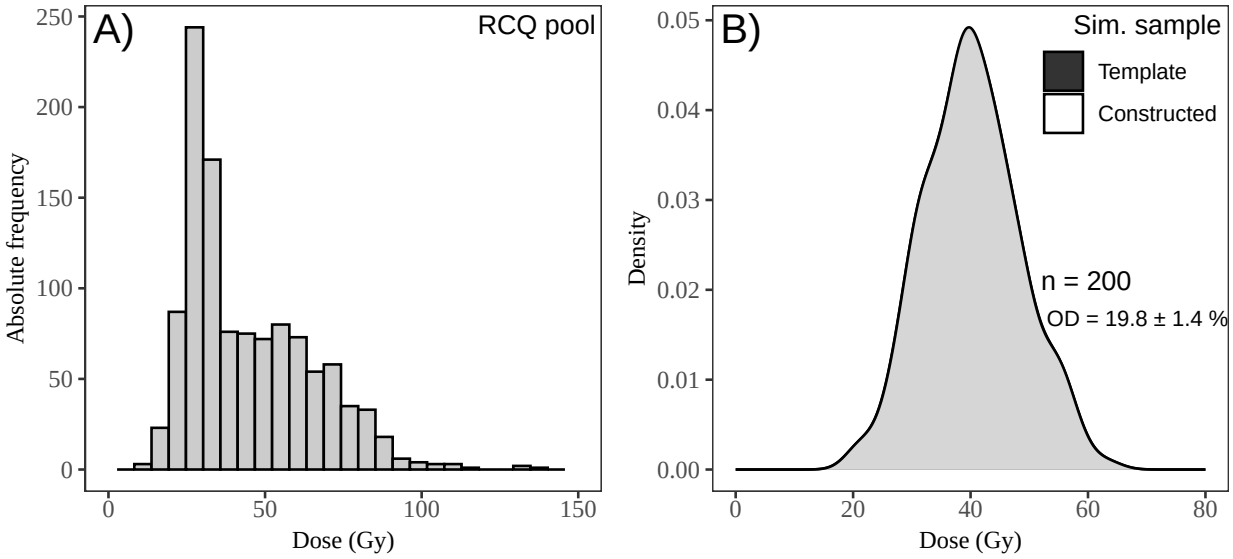

Figure 2.3: Dose distributions of a simulated single-grain (SG) quartz data set. (A): Total pool of Risø Calibration Quartz (RCQ) grains (B): Density plot of a simulated 200 grain SG sample with an OD of $19.8 \pm 1.4\%$ overlain the lognormal template used to simulate the dataset.

2.3 Exploratory survey

In practice, it is a good idea for a BayLum user to assess MCMC chain convergence, not only from Rubin-Gelman scores, but from visual inspection of trace-plots as well. In this chapter, it was often not feasible to make visual evaluations of trace-plots, with hundreds - and sometimes thousands, BayLum runs. Therefore, in this chapter, I define convergence solely as reaching a 95% confidence interval upper limit Rubin-Gelman score below 1.050.

2.3.1 BayLum and the number of MCMC samples to reach convergence

In many cases we will want to incorporate several OSL samples into a single BayLum model. This is for example required in order to model systemic errors or in order to impose stratigraphical constraints between samples. But will the number of OSL samples have an effect on the number of MCMC samples required to reach convergence with BayLum? In the following, I used BayLum to derive equivalent doses for a 3-sample and 6-sample model in which all samples where duplicates of the 23-aliquot multi-grain sample 207733. My goal was to record the number of MCMC samples required for BayLum to reach convergence - which I here define as obtaining a upper limit Rubin-Gelman diagnostic score below 1.050 for "A" (age),

"D" (dose) and "sD" (uncertainty on D) BayLum parameters. To achieve this goal, I specified a relatively low number of initial MCMC iterations for BayLum to run for (so that convergence would not be reached), and then I continuously extended un-converged BayLum-models with the same number of MCMC samples each run until convergence was eventually reached (see chapter 3). For the 3-sample model, I set an initial number of 150k MCMC samples ("adapt" = 50k, "burnin" = 50k, "iter" = 50k). For the 6-sample model, I specified 350k MCMC samples each run ("adapt" = 50k, "burnin" = 150k, "iter" = 150k). Finally, I compared these two sample setups with a third 6-sample setup built from duplicates of the 344 grain single-grain 207749 sample ("adapt" = 50k, "burnin" = 12.5k, "iter" = 12.5k). Each setup was run 8 times. With each BayLum run, I specified a "lognormal_A" dose dispersion model, and a single saturating exponential (going through the origin) DRC fit. I show the results of this experiment in Figure 2.4.

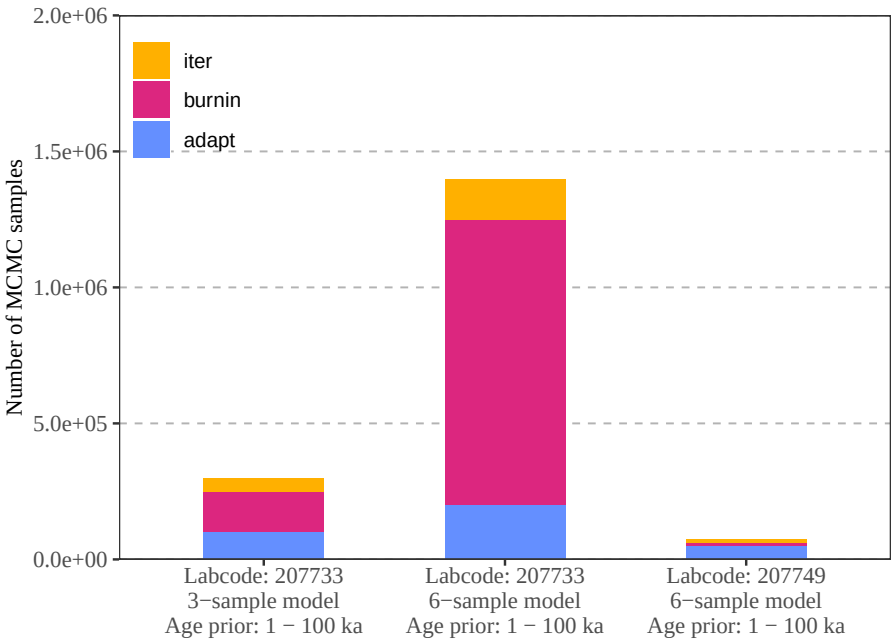

Figure 2.4: Median number of BayLum MCMC samples (sum of "adapt", "burn" and "iter" arguments) required to achieve convergence for three different sample setups. Each median summarize 8 replicate BayLum runs. The three sample setups are: a 3-sample setup from multi-grain sample 207733 (meaning that the setup contains three identical samples - all of them 207733), a 6-sample setup from multi-grain sample 207733 and a 6-sample setup from single-grain sample 207749. I used an age prior of between 1 and 100 ka for all BayLum runs. Convergence is here defined as achieving a Rubin and Gelman diagnostic below 1.050 for "A", "D" and "sD" model parameters.

Figure 2.4 reveals that in the case of sample 207733, the number of OSL samples run with BayLum greatly affects the amount of MCMC samples required to reach

convergence. Going from a 3-sample to 6-sample setup required about four times the median number of MCMC iterations for BayLum to reach convergence. Strikingly, the 6-sample 207749 setup required by far the fewest number of MCMC samples of the three. This is likely, from experience, a result of the fact that far more aliquots/grains are evaluated per sample in 207749 (23 vs 344). I am however unable to point towards any particular mechanism that would explain why this is.

It is possible with BayLum to specify an age prior - that is, it is possible to provide BayLum with prior information about the age of a sample. This could potentially speed up convergence, because it may be providing the MCMC sampler with better initial starting values. During the course of the previous experiment, I became quite familiar with what dose and age that BayLum should eventually converge on for sample 207733 - an age at about 20-40 ka. I now provide BayLum with this prior age information and rerun 207733 (8 replicate runs). Figure 2.5 shows the median number of MCMC samples required to reach convergence for the 6-sample 207733 setup described in the beginning of this section, using either an "AgePrior" of 1-100 ka or 20-40 ka. The method to estimate this number of MCMC samples mirrors the method described in the beginning of this section.

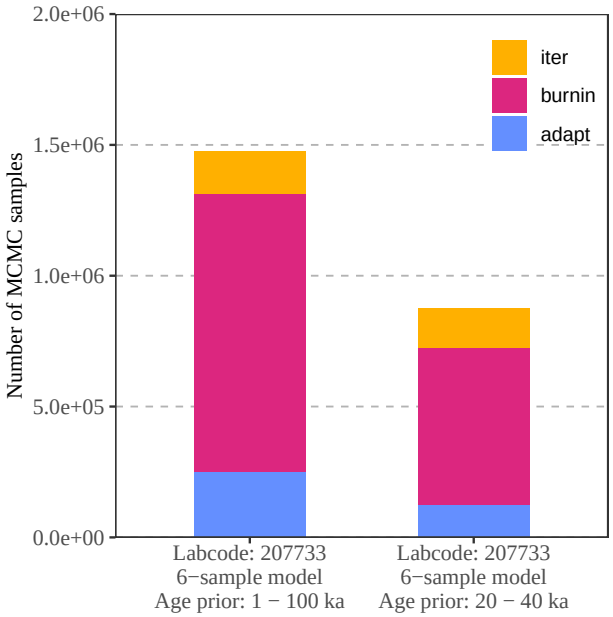

Figure 2.5: Median number of BayLum MCMC samples (sum of "adapt", "burn" and "iter" arguments) required to achieve convergence using either an age prior of 1-100 ka or an age prior of 20-40 ka. We use the same sample setup for each prior treatment, each replicated 8 times. BayLum is run on a 6-sample setup from multi-grain sample 207733 (meaning that the model is built from six samples - all of them sample 207733). Convergence is here defined as achieving a Rubin and Gelman diagnostic below 1.050 for "A", "D" and "sD" model parameters.

Figure 2.5 appears to support the notion that better age-priors reduces the number of MCMC samples required to reach convergence. For when using an age-prior of 20-40 ka, the median number of MCMC samples required for convergence was little more than half compared with using a 1-100 ka age-prior.

These results highlight, that if several OSL samples are required to be run within a single BayLum model, it may be worthwile to measure more aliquots/grains per sample and/or specify suitable age-priors. But what are good age-priors? By running an exploratory BayLum-run not meant to converge, it may be possible to extract information about the dose regions in which samples will fall. However, a word of warning: when aggressively optimizing the age-prior, I would recommend to survey results and see if any sample appears to hit a "boundary" - for example: an estimate of 49.2 ka, 68% credible interval [42.3 ka, 50.0 ka] has an upper credible interval limit near its highest density peak - a telltale sign that the model may have run into what I term an "age-prior boundary".

2.3.2 BayLum and the effect of extreme dose measurements

Single-grain dose distributions sometimes reveal the presence of an observation (or more) so extreme "...as to arouse suspicions that it was generated by a different mechanism." (Hawkins, 1980). These extreme dose observations could both be doses that are many times larger than the main body of the data, or doses consistent with zero. We might wonder how these grains affect equivalent dose estimation. In the following, I added an increasing number of a zero-dose grain with a measured dose of 0.4 ± 2.0 Gy. The results of adding this zero dose grain are seen in Figure 2.6A and 2.6B.

Intuition predicts that by adding zero-dose grains, equivalent dose estimates will decrease. But it appears that BayLum (with a "lognormal_A" dose dispersion model) behave contrary to this notion, when in this case more than 34 zero-dose grains were added (Figure 2.6). However, I find that zero-dose outlier effects depend on the nature of the zero-grains involved. In my original experiment (not shown), I included a 0.3 ± 0.2 Gy grain instead. Here, also ADM had a strong upwards reaction. Below an addition of 10%, BayLum (lognormal_A) does seem initially resistant, with normalized estimates (normalized to no-outlier estimates) around 1. The same is true for BayLum (cauchy) and BayLum (gaussian). BayLum (lognormal_M) appears affected, even with few zero-dose additions.

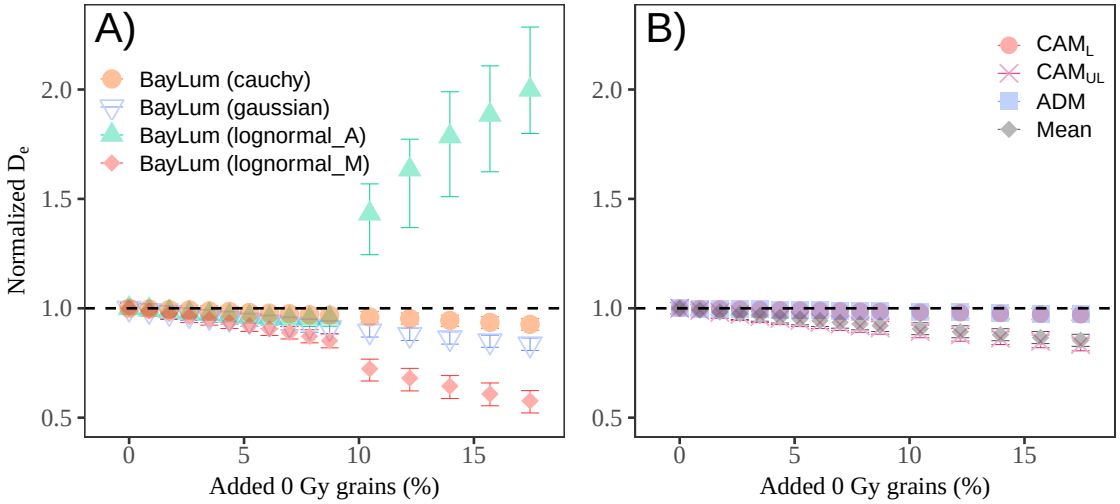

Figure 2.6: Effect of outliers on equivalent dose estimation. (A): D_e as a function of the number of 0 Gy grains added to sample 207749 (in percent of the original number of grains: 344) using the various BayLum dose dispersion models. (B): As (A), but for the CAM (logged and unlogged), ADM and the arithmetic average (Mean). Equivalent dose estimates have been normalized to the BayLum equivalent dose estimate (for each dose-dispersion model) obtained when 0 outliers were added to sample 207749.

I repeated the zero-dose outlier experiment, only this time adding repeats of a 199 ± 10 Gy grain (Figure 2.7A and 2.7B). I then estimated the resulting equivalent doses using BayLum, CAM (both logged and unlogged) and ADM ($\sigma_m = 0.30$). With each BayLum run, I specified a DRC fit by a single saturating exponential going through the origin. The results can be seen in Figure 2.7.

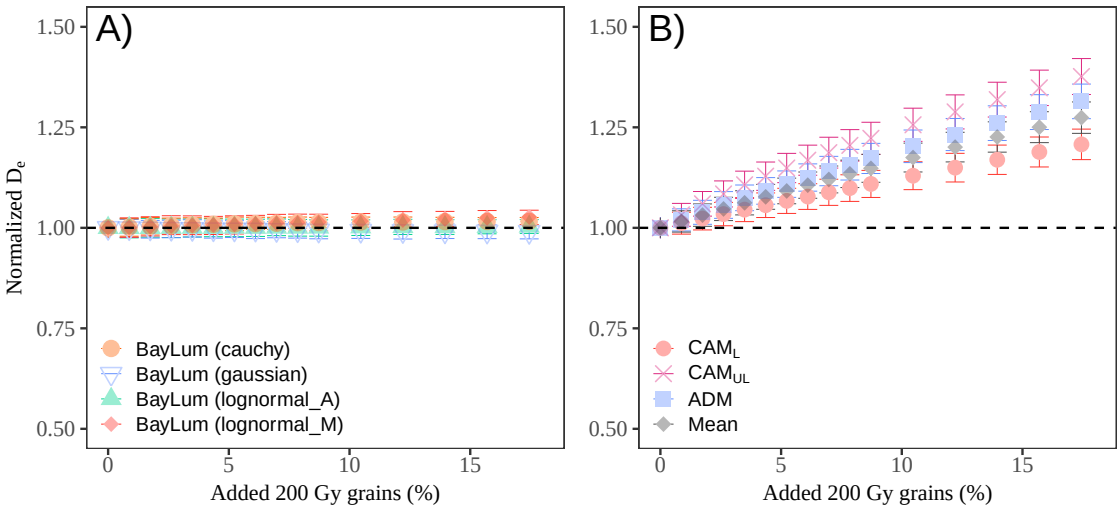

Figure 2.7: Effect of outliers on equivalent dose estimation. (A): D_e as a function of the number of 0 Gy grains added to sample 207749 (in percent of the original number of grains: 344). (B): D_e as a function of the number of 200 Gy grains (in percent) added to sample 207749. Equivalent dose estimates have been normalized to the BayLum equivalent dose estimate (for each dose-dispersion model) obtained when 0 outliers were added to sample 207749. ,

It appears that BayLum and all its dose-dispersion variants are resistant to high dose outliers over the full tested range (Figure 2.7A). The same cannot be said for CAM, ADM and the basic mean, all of whom produce increased estimates with even the fewest tested high-dose outlier additions tested. The estimates then increase continuously with more high-dose outliers added.

2.3.3 Must all BayLum parameters converge?

When using MCMC in Bayesian inference, it is crucial that the sampled marginal posteriors are stationary - which is to say that the location and shape parameters of the distributions will not change if additional MCMC samples are drawn. BayLum provides convergence diagnostic capabilities for three parameters of the BayLum model (A, D and sD) out of the box. But in reality, the hierarchical model within BayLum contains many more parameters. When running BayLum using a single saturating exponential to fit DRCs, at least three additional parameters directly related to the fitting of the DRC are created for each aliquot/grain. A 100-grain sample will have 300 of these DRC parameters. The parameters I am referring to are $a_1, a_2, ..., a_{100}$ and $b_1, b_2, ..., b_{100}$ and $sigmaf_1, sigmaf_2, ..., sigmaf_{100}$, where $sigmaf_i$ represents the goodness-of-fit of the i^{th} DRC fit, while a_i and b_i are perhaps better known as the "A" and "D_c" parameters of the following function:

$$L_x/T_x = A \times (1 - exp(D/D_c))$$

By modifying a BayLum modelling R-function, "Age_Computation()" (Christophe et al., 2023)), I was able to run BayLum repeatedly for 207749, setting a variety of MCMC iterations to run for. Meanwhile, I assessed convergence for these DRC parameters not normally given in the BayLum output. As it turns out, reaching full model convergence was not an easy accomplishment. In Figure 2.8, I show the percentage of DRC-parameters (a, b and $sigmaf$) which reached convergence as a function of the number of MCMC samples drawn. For all observations shown, A, D and sD parameters all had upper limit Rubin-Gelman scores below 1.050.

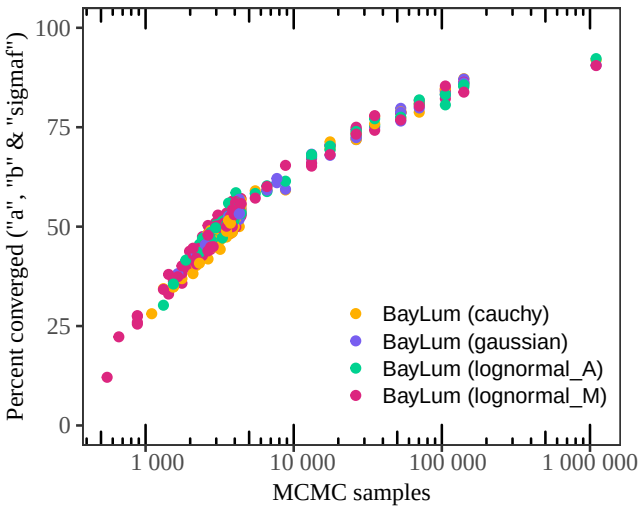

Figure 2.8: Percent of a, b and $sigmaf$ parameters which achieve upper limit Rubin-Gelman diagnostic scores below 1.050 as a function of total MCMC samples run in BayLum (n = 463) using sample 207749. For each observation, the A, D and sD parameters all have upper limit Rubin-Gelman scores below 1.050.

Figure 2.8 illustrates clearly, that even when A, D and sD parameters all reach convergence, it does not necessarily mean full convergence of the DRC-parameters. In fact, it appears possible to have A, D and sD marginal posteriors be stationary whilst almost 90% of DRC-parameters have upper limit Rubin-Gelman scores above 1.050. Also, even when 1 million MCMC samples is run for a single sample, less than 100% (≈90%) of DRC parameters converged.

Of course, the limit below which I define convergence is an arbitrary one (Rubin-Gelman score < 1.050). And it could be argued that the DRC parameters which did not converge in Figure 2.8 are right on the edge of this potentially overly strict limit. But this is not quite the case. Figure 2.9 illustrates that the median Rubin-Gelman score of DRC parameters that did not converge in Figure 2.8 for each dose model, for each BayLum-run, is well above 1.050. With less than 10,000 MCMC samples, the medians hover between Rubin-Gelman scores of 1.20 and 1.45.

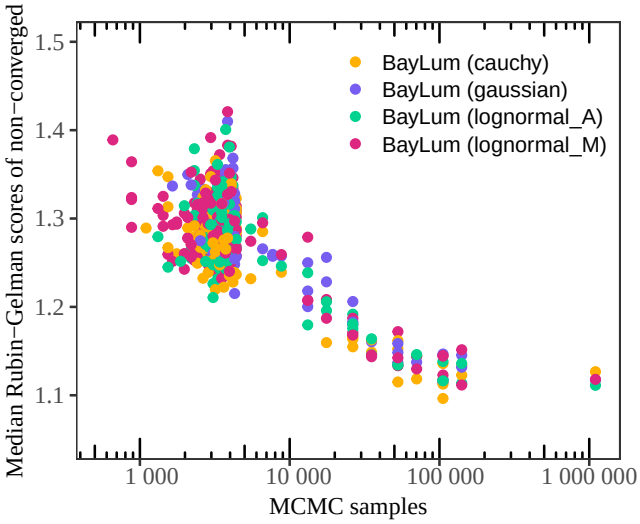

Figure 2.9: Median Rubin-Gelman scores for DRC parameters (a, b and $sigmaf$) that did not converge as a function of total MCMC samples run in BayLum (number of runs = 463) using sample 207749. For each observation, A, D and sD parameters all have upper limit Rubin-Gelman scores below 1.050.

The question then becomes: does it matter for central dose estimation in this hierarchical model, that bottom level DRC-parameters do not converge? Figure 2.10 shows BayLum D_e estimates vs the percentage of DRC-parameters which reached upper limit Rubin-Gelman scores below 1.050 (while the A, D and sD parameters all scored below 1.050). These results appear to indicate that BayLum estimates are uncorrelated with the percentage of DRC parameters achieving upper limit Rubin-Gelman scores below 1.050 (when A, D and sD did converge).

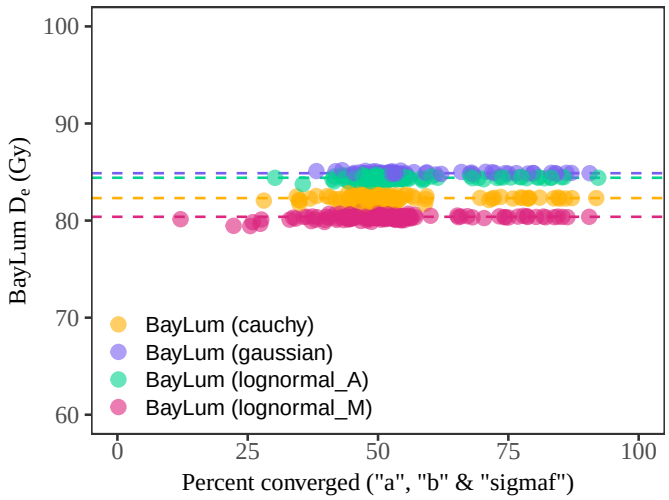

Figure 2.10: BayLum D_e (Gy) as a function of the percent of a, b and $sigmaf$ parameters who achieve upper limit Rubin-Gelman diagnostic scores below 1.050 (n = 463). For each observation, A, D and sD parameters all have upper limit Rubin-Gelman scores below 1.050. Dashed horizontal lines (slope = 0) intercept the y-axis at the value of the measurement with the highest percent convergence of DRC parameters for each dose dispersion model

These results support the practice of judging convergence only from the A, D and sD parameters.

2.3.4 BayLum and sample size

When we sample sediment for the purpose of OSL dating, our hope is that each of our samples are representative of the sediment we wish to make inferences about. We can make a similar statement about the preparation of aliquots or SG discs from the total pool of extracted quartz grains; we hope that the grains used to populate aliquots or SG discs are representative of the total pool of extracted quartz grains. How many grains are needed for BayLum such that the dose estimate variation between a sample and the total population of extracted quartz grains remains acceptably small?

In this section, I treated single-grain measurements of sample 207749 (n = 344, OD $\sim$ 40%), and sample 981009 (n = 300, OD $\sim$ 62%) as if these dose distributions were the total dose distributions of extracted quartz grains for two hypothetical OSL samples. I then randomly drew a variable number of grains to imitate the process of preparing SG discs. It is important that each random draw occurs with replacement, such that the dose population (and thus its characteristics) which is drawn from, is unchanged after each draw. I refer to these characteristics, or location and shape parameters, as the "true" parameters for these hypothetical OSL samples. For each sample size,

I drew 8 replicate samples and determined BayLum D_e values for each dispersion model (using a single-saturating exponential function forced through the origin). I also calculated CAM (both logged and unlogged) and ADM ($sigma_m = 0.30$) dose estimates. I repeated this experiment for the simulated sample (n = 200, OD ∼ 20%). Only converged results were used in analysis.

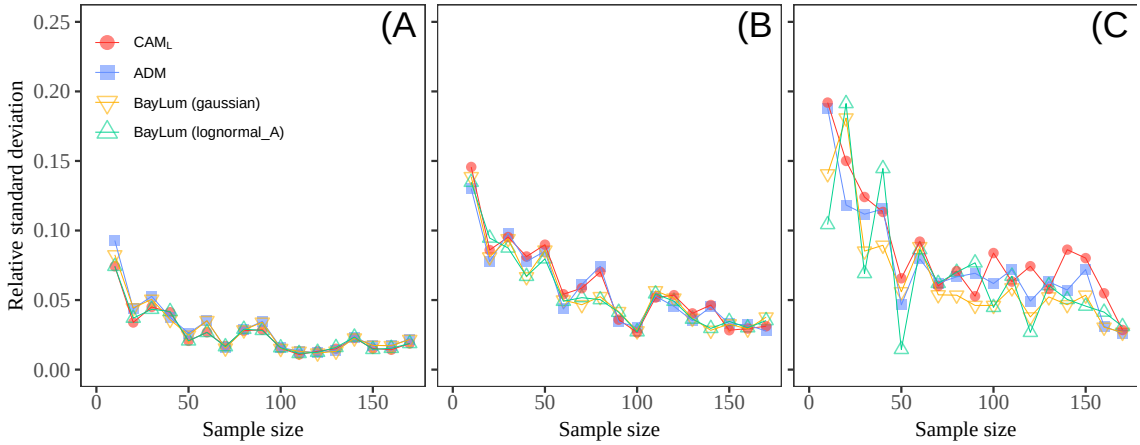

Figure 2.11: Relative standard deviation of replicate BayLum runs (per sample size per dose dispersion model). (A): Simulated sample, OD ∼20%. (B): 207749, OD ∼40%. (C): 981009, OD ∼62%

Figure 2.11 shows the relative standard deviation (RSD) calculated from eight equivalent dose estimates for each sample size (per dose model). In terms of how prone dose model estimates are to vary as a result of random sampling, it is difficult to argue that one dose model is less or more sensitive to sample size. But it does seem as though over-dispersion affects the RSD. Through the full range of sample sizes, the normalized mean of replicate samples per dose model is consistent with unity within 1σ (data not shown) - sample estimates were normalized to the "true" estimates of each hypothetical OSL sample of chemically treated grains. When OD of the sample is low (∼20%, Figure 2.11A), the standard deviation of 8 replicate sample estimates was about 5% when using a sample size of 20 grains. For the sample with an OD of 40%, RSD dropped below 5% only when more than 50 grains were used (Figure 2.11B). When the sample OD is ∼60% (Figure 2.11C), it is not clear that the RSD ever consistently dropped to below 5%. Rather, RSD fluctuated around 5% from 50 grains and up.

In the previous paragraph, I compared results only according to when an RSD of 5% was achieved for any of the OSL samples. However, although the 5% RSD limit

is acceptably small is is also arbitrary. Thus, if a RSD of say 10% is acceptable, then samples size is less important.

2.4 Added feature: *write_BayLumFiles()*

BayLum modelling functions require the input of the "DataFile"-object. To generate this object, the "BayLum" R-package comes with functions (e.g., "Generate_DataFile_MG()") that expect a set of files detailing which aliquot/grains should be included, what the summation intervals should be, the environmental dose rate, etc.

Setting up these files, especially for a large amount of samples, can become a tedious expenditure of time. And worse still, manual manipulation of .csv-files are bound to produce errors, which unfortunately are often difficult to locate based on the error-messages provided by "R". This is not to mention, that it is sometimes necessary to make changes to our initial setup. For example, perhaps a different set of grains is to be analyzed or perhaps our estimate of the environmental dose rate changes. To overcome these obstacles while working with 40 samples from Kostenki 17, I wrote the function "write_BayLumFiles()" - now available with BayLum v.0.3.1 (Christophe et al., 2023). This function takes as input all the relevant information to create the necessary .csv-files. This guarantees typo-free files to the specifications of BayLum. This also brings with it the possibility for the user to check the history of inputs and to automate experiments that require changes to be made to the .csv-file information in between runs. I especially found it useful to output Analyst results into a combined .csv-file for all samples. With a few added columns (such as source dose rate), it is possible to automatically create all folders and all .csv-files with one single function call. The exception are the binx-files, which must be placed inside each folder manually.

Below I describe one example of how to use "Write_BayLumFiles":
In the example given in Listing 2.1, two sample folders ("Sample1" and "Sample2") will be created inside the already existing folder "BayLum_files/SiteA". "Sample2" will have two sub-folders, one for each binfile (specified by "BinPerSample"). Aliquot positions 1-3 will be written to "Sample1", 9-11 to "Sample2/1" and 17-20 to "Sample2/2". Environmental dose rate ("DRenv") and its absolute uncertainty ("DRenv.error") will be written in the same order to each of the sample folders (i.e.,

1.28 will be written to the first created folder). When only 1 number is input (e.g., "DRsource" in this example), all sample folders will have this value stored within the relevant .csv-file.

```
write_BayLumFiles(
  folder = "BayLum_files/SiteA",
  SampleNames = c("Sample1","Sample2"),
  DiscPos = list(data.frame("position" = c(1,2,3)),
                 data.frame("position" = c(9,10,11)),
                 data.frame("position" = c(17,19,20))),
  BinPerSample = c(1,2),
  DRenv = c(1.28, 1.72, 1.72),
  DRenv.error = c(0.06, 0.07, 0.07),
  DRsource = 0.808,
  DRsource.error = 0.001,
  signal.integral.min = 6,
  signal.integral.max = 10,
  background.integral.min = 346,
  background.integral.max = 395,
  nbOfLastCycleToRemove = c(2,2,2)
)
```

Listing 2.1: R-function within BayLum (Christophe et al., 2023) to write the csv-files required by BayLum modelling functions.

2.5 Future BayLum perspectives

Through my work with the "BayLum" R-package, both as a user and as a developer, I have observed several ways to improve both convergence runtime and user experience of the BayLum "R-package". Below, I highlight three possible avenues of development, which I believe will make significant improvements to BayLum for the general user, but that I unfortunately did not have time to assist in implementing.

- Incorporate automatic selection of plausible initial values for the MCMC. Runtime until convergence may dramatically improve if the MCMC is given a good place to start.
- Generate_DataFile() (Christophe et al., 2023) is a function that extracts the neccessary information from binx-files. Any imperfection within the binx-file, and the BayLum-function is likely to fail - either outright or produce

incoherent values. It is my belief that the "Generate_DataFile()" can be re-written to become much more flexible over minor binx-file imperfections. An imperfection could for example be that the ordering of regenerative doses is not the same for all aliquots to be modelled inside the binx-file.

- Make error-messages meaningful. Currently, many error-messages that a user is likely to experience will be generic "R"- messages and not specific to what possibly went wrong within the context of BayLum. This complicates troubleshooting alot.

3 Reducing computation time in the R-package 'BayLum'

Frederik Baumgarten[1], Anne Philippe[2], Guillaume Guérin[3], Sebastian Kreutzer[4,5]

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

[2] Laboratoire de Mathématiques Jean Leray, Université de Nantes 2 rue de la Houssinière, BP 92208 44322 Nantes Cedex 3, France

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

[4] Institute of Geography, Ruprecht-Karl University of Heidelberg, 69120 Heidelberg, Germany

[5] Archéosciences Bordeaux, UMR 6034, CNRS-Université Bordeaux Montaigne, Pessac, France

Chapter outline

Here I present recent features implemented into the BayLum R-package. These features primarily seek to improve the required runtime of BayLum. This work has been published in *Ancient TL* (Baumgarten et al., 2023).

Abstract

'BayLum' is an R-package that facilitates the application of Bayesian models to the field of OSL dating. Here we present two recent feature updates to 'BayLum', significantly reducing computation time and improving general use. The first feature allows users to parallelize the computations involved in the MCMC sampling of values, while the second introduces the ability to extend a 'BayLum' model, which has run to completion without converging. All updates are automatically available with 'BayLum' v0.3.1.

Keywords: Age model, Chronology, MCMC algorithm, Luminescence dating, OSL

3.1 Introduction

'BayLum' is an R – package (R Core Team, 2023) that gives users the tools to easily apply the Bayesian models presented in Combès et al. (2015) and Combès and Philippe (2017) to luminescence dating data. See, for example, the work of Heydari et al. (2020), where an OSL chronology is provided for the paleolithic site of Mirak, Iran, using 'BayLum'. In this work, they showed that the age uncertainty can be reduced significantly by imposing stratigraphic order – a feature of 'BayLum'. Since the introduction of 'BayLum' (Philippe et al., 2019), 'BayLum' has grown by drawing resources from the ever-developing R-landscape around it. The latest iteration of 'BayLum' (v0.3.1) (Christophe et al., 2023) now employs `` `runjags' `` (Denwood, 2016) as the R to JAGS (Plummer, 2003) facilitator, which has made possible two key features of 'runjags' to be used inside 'BayLum': (i) MCMC-sampling parallelization and (ii) the ability to extend a model (drawing additional MCMC samples after a model has already run to completion). This paper will highlight these two new features of 'BayLum' and show examples of how to use them.

3.2 Problem: Stationary distributions require long run times

The Bayesian models produced with 'BayLum' infer parameter estimates (such as equivalent dose and age) from marginal posterior distributions of these parameters. This is to say that 'BayLum' takes the output of the Bayesian approach, a posterior

distribution, and evaluates the dimensions of individual variables. 'BayLum' constructs these distributions via Markov Chain Monte Carlo sampling. The result of the MCMC sampling is a chain of values, each link consisting of a combination of values from all parameters in the Bayesian model. A distribution can then be constructed for each parameter, given its value in each link. To let the MCMC converge on the solution, we skip a number of the first iterations (burn-in phase) and only then begin constructing the distributions. To be confident in the results, the distributions must be stationary – that is, the location and shape of each distribution must not change if we draw additional samples. 'BayLum' assesses if distributions are stationary and independent of initialization of the MCMC by constructing multiple chains instead of one. If the distributions from each chain agree with each other, we can be confident that the chains have converged to a single solution. By default, 'BayLum' uses three MCMC chains – a suitable balance between the power to detect non-convergence and the computational resources required (the number of chains is fully customizable by the user). 'BayLum' formalizes the question of convergence by incorporating as output the Rubin and Gelman diagnostic (Gelman and Rubin, 1992), which compares within-chain and between-chain variance. A common rule of thumb is that the upper 95 % credible interval limit of this diagnostic value indicates convergence when below 1.05.

For many practical applications of OSL dating, the number of iterations (or links in each chain) required to reach convergence is high (>500 000) – and higher still when 'BayLum' models incorporate many OSL samples as is the case with high-resolution chronologies. Because MCMC chains are to be processed consecutively, the overall process can become very time-consuming. For example, using a computer equipped with a 11^{th} Gen i7-1185G7 clocking at 3.0 GHz (which has a relatively high single-core threading performance rating), runtimes can extend beyond several days. Furthermore, even when a model completes, not all of the model's parameters may have converged – a result which could require a complete re-run of the 'BayLum' modelling function.

3.3 'BayLum' feature: MCMC parallelization

Previous versions of 'BayLum' could only process MCMC chains consecutively using a single processor core. With parallelization, it is now possible to assign n chains

out onto n CPU processor cores. This allows each chain to be processed concurrently, and the runtime will (ideally) approach 1/n when compared to the time for running n chains using a single core. We tested this using 'BayLum' models where OSL example sets GDB3 and GDB5 were used (both included with the 'BayLum' package) to produce 2-sample models. Figure 1A shows that when running 4 000 total iterations per chain, we see a significant runtime reduction when running the model using parallelization (`jags_method = "rjparallel"`) as compared to using only a single CPU core (`jags_method = "rjags"`). Reduction increases with the number of MCMC chains constructed in the model, which is what we expect. We observed a reduction of 65 % for a 3-chain setup and 72 % for a 4-chain setup. The minor differences we see from the theoretical 1/n-rule most likely arise from runtime inside the 'BayLum' model functions, which is not due to the iteration of MCMC sampling. We also see from Figure 1B that this reduction is consistent with increasing numbers of iterations. Example 1 (Sec. 3.3.1) shows how to apply parallelization in 'BayLum' v0.3.1. Note that our model testing was carried out using the High-Performance Computing Cluster "Sophia" (Technical University of Denmark, 2019). The same code run on a desktop PC will show the same relative reductions but may show poorer runtimes, not only because of lower overall computation power but also - and more likely - due to advanced power throttling measures of modern CPU architectures implemented to prevent overheating in prolonged high-load situations.

3.3.1 Example 1

In the example below (Listing 3.1), which we kept as simple and user-friendly as possible, we show how to achieve parallelization. The key argument to set is `jags_method = "rjparallel"`. We use the example data included within 'BayLum' at installation.

```
# MCMC parallelization example ####
# load libraries
library(BayLum)

# load example DataFiles GDB3 and GDB5
data(DATA1)
data(DATA2)

# combine DataFiles
# (we now have a 2-sample DataFile)
```

```
DF <- combine_DataFiles(DATA1, DATA2)

# construct BayLum model
BayLum_model <- AgeS_Computation(
  DATA = DF,
  SampleNames = c("GDB3", "GDB5"),
  Nb_sample = 2,
  BinPerSample = c(1, 1),
  LIN_fit = FALSE,
  Origin_fit = TRUE,
  Iter = 1e+03,
  burnin = 5e+02,
  adapt = 5e+02,
  n.chains = 3,
  jags_method = "rjparallel"
)
```

Listing 3.1: R Code: Achieving parallelization

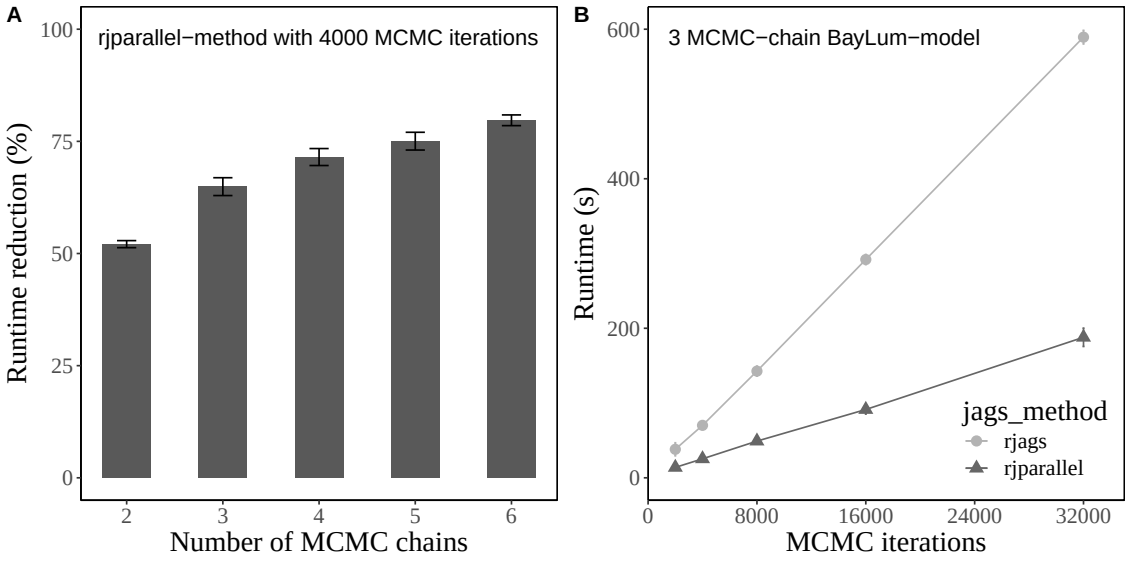

Figure 3.1: (A): Runtime reduction in percentage when running a 'BayLum' model with fixed iterations vs a varying number of MCMC chains using GDB3 and GDB5 example sets included within 'BayLum'. (B): Runtime in seconds vs the number of MCMC iterations for a 3-chain 'BayLum' model also using GDB3 and GDB5. All estimates show mean ± sd (n=8). To run the model, we used the High-Performance Computing cluster named "Sophia" owned by DTU. Arguments `"rjags"` and `"rjparallel"` entail whether 'BayLum' is run using a single CPU core (`'rjags'`) or run in parallel on several cores (`'rjparallel'`).

3.4 'BayLum' feature: extend the 'BayLum' model

Unfortunately, 'BayLum' model chains will not always converge within the specified number of iterations. In previous versions of 'BayLum', the 'BayLum'-model would likely need to run again with a higher number of iterations. The added runtime of re-running 'BayLum' can now be avoided by extending the non-converged model instead of building it again from scratch. In this case, all non-converged model iterations are treated as burn-in. See Example 2 (Sec. 3.4.1) for an illustration of how to extend 'BayLum' models.

3.4.1 Example 2

In Example 1 (Sec. 3.3.1), a model was built to show how parallelization could be achieved. The Rubin and Gelman convergence diagnostics from that model reveal evidence that not all MCMC chains converged (see `"D (Dose)"` for GDB5, Table 3.1).

Table 3.1: Rubin and Gelman convergence diagnostics for three parameters of the 'BayLum'-model in Example 1. We show only the upper 95 % credible interval limit.

Sample	A (Age)	D (Dose)	sD (Stand. deviation)
GDB3	1.006	1.022	1.004
GDB5	1.007	1.065	1.000

However, we can now add iterations to the 'BayLum'-model in order to achieve convergence (Listing 3.2):

```
# extend MCMC sampling of BayLum-model
BayLum_model_extended <- AgeS_Computation(
  DATA = BayLum_model,
  SampleNames = c("GDB3", "GDB5"),
  Nb_sample = 2,
  BinPerSample = c(1, 1),
  LIN_fit = FALSE,
  Origin_fit = TRUE,
  Iter = 1e+04,
  burnin = 0,
  adapt = 5e02,
  jags_method = "rjparallel"
)
```

Listing 3.2: R Code: Extending model

Rubin and Gelman's convergence diagnostics now show we can be confident about all the parameters (Table 3.2).

Table 3.2: Rubin and Gelman convergence diagnostics for three parameters of the 'BayLum' model from example 1 (Sec. 3.3.1). We show only the upper 95 % credible interval limit.

Sample	A (Age)	D (Dose)	sD (Stand. deviation)
GDB3	1.002	1.007	1.000
GDB5	1.001	1.010	1.004

3.5 Conclusions

In this report, we introduced two feature updates to the R-package 'BayLum'. Together, they allow users to parallelize MCMC sampling and extend BayLum-models - both features significantly reduce the time needed to build a viable 'BayLum'-model.

Acknowledgments

We thank Geoff Duller for his thorough and supportive comments. We also gratefully acknowledge the computational and data resources the Sophia HPC Cluster provided at the Technical University of Denmark, DOI: https://doi.org/10.57940/FAFC-6M81.

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

Frederik H. Baumgarten[1*], Kristina J. Thomsen[1], Guillaume Guérin[2], Jan-Pieter Buylaert[1], Andrew S. Murray[1,3]

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

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

[3] Nordic Laboratory for Luminescence Dating, Department of Geoscience, Aarhus University, and DTU Physics

*Corresponding Author: fhaba@dtu.dk

Chapter outline

Chapter 4 provides a test of accuracy for single-grain OSL dating of Eemian age samples, with expected equivalent doses in the 100 - 200 Gy dose range. Results of conventional single-grain dating methodology are compared with both multi-grain OSL dating of quartz and newer single-grain methodology. This work was presented as an oral presentation in the LED2023 conference (June, 2023) and has been submitted for publication in a reduced form to the LED2023 special edition of *Quaternary Geochronology* (Baumgarten et al., submitted).

The version of the article presented here is a slightly modified version of the one submitted. I provide a full changelog at the end of this chapter.

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, very little testing of the accuracy of single-grain dating has been published. 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 Eemian age samples (∼128 ka). We show that standard multi-grain quartz dose estimation results in dose estimates in good agreement with the predicted doses (average ratio of 1.00 ± 0.04, n = 5), but that standard frequentist single-grain dating procedures significantly underestimate the age controls, e.g., the mean CAM-to-predicted dose-ratio is 0.56 ± 0.08 (n=5). Using the D_c criterion to reduce the effect of saturated grains, we were able to improve the accuracy of our dose estimates to 0.74 ± 0.05 (n = 5), which however still underestimates the age control significantly. The most accurate single-grain dose estimation is obtained using Bayesian analysis ("BayLum", ratio of 0.95 ± 0.08, n = 5). 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

4.1 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 and Roberts, 2011; Jacobs 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., 2011, 2013; Demuro et al., 2019, 2023), or because they can be identified as sources of contamination from e.g., post-depositional mixing (e.g., Roberts et al., 1999; Feathers et al., 2006; Jacobs et al., 2011; 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 CAM age to the reference age was 0.90 ± 0.02 (using a standard frequentist approach), indicating an average underestimation of 10%. 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. However, single-grain quartz analysis have also been found to be in agreement with independent age control 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 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 SAR dating for five Eemian samples with known absorbed doses in the 100-200 Gy range. Our primary questions are (i) how do the Central Age Model (CAM, Galbraith et al., 1999) and the Average Dose Model (ADM, Guérin et al., 2017) perform on these samples? And (ii) do rejection criteria improve equivalent dose estimation? We also test and compare the performance of BayLum (Combès et al.,

2015; Combès and Philippe, 2017; Christophe et al., 2023), which takes a hierarchical Bayesian approach to equivalent dose and age estimation. Here, the usual steps from 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).

4.2 Samples and independent age control

For the purpose of testing single-grain OSL dating of quartz in the 100-200 Gy dose range, we chose five Eemian samples for which the burial age is constrained around 128 ka. By this time, Earth was in the early phase of the last interglacial (Marine Isotope Stage 5e) - an interglacial warm period that correlates with the Eemian of Northwest Europe, and is characterized by a warm climate and high global sea levels (Kaspar et al., 2005; Kopp et al., 2009).

Three of our five Eemian samples (Sample codes: 981007, 981009, 981013) were collected in 1998 at Gammelmark, Denmark, 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 obtained OSL samples from the Brunsbjerg section of the Gammelmark cliffs located on the south east coast of the Jutland peninsula and describe a sequence of marine sediments in upwards succession of a marine transgression layer (Figure 4.1).

Among these layers are Anadonta sand and Cyprina clay layers for which pollen analysis showed that they have been deposited within the first couple of centuries of the Eemian (Funder et al., 2002). The marine succession culminates in Tapes sand, named so for the shallow-water mollusc fossils found within the marine deposits. Funder et al. (2002) place the timing of the regression of marine sedimentation to around the interglacial highstand at 128 $\pm$ 1 ka (McCulloch and Esat, 2000) with marine sedimentation ceasing within a few thousand years, capping the end of marine sedimentation to around 125 ka (Menke, 1985; Kristensen et al., 2000; Funder et al., 2002; Murray and Funder, 2003). They then define an age for the whole marine section of between 133 and 125 ka. Samples 981007 and 981009 were collected from the layer of Tapes sand, while 981013 was collected from the underlying tapes sand to Cyprina clay transition-layer. Buylaert et al. (2011) revised the quartz

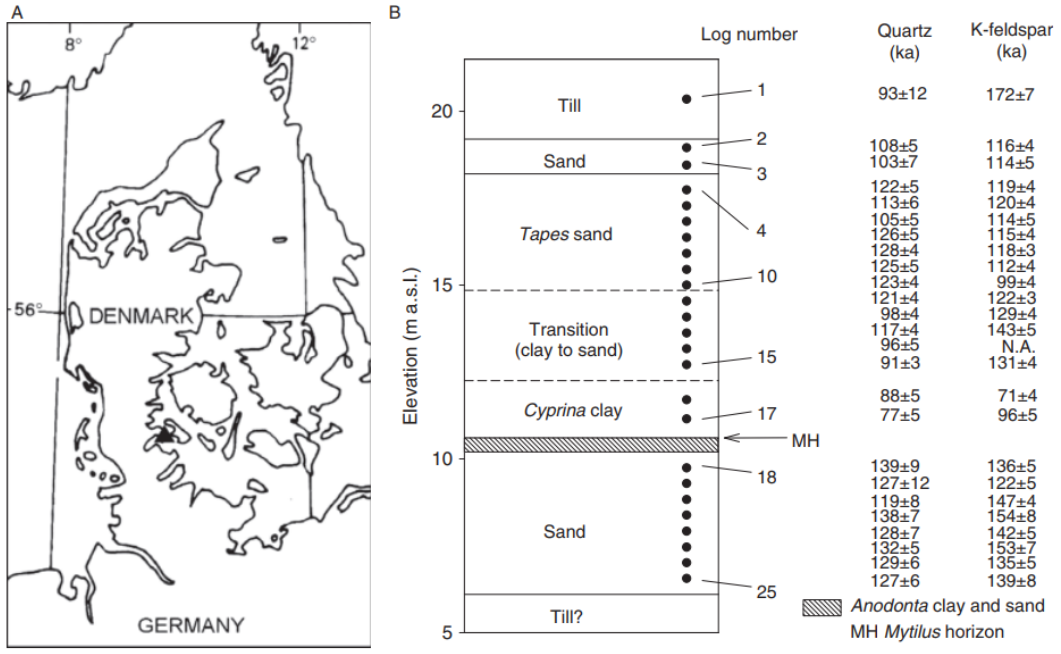

Figure 4.1: (A): Map of Denmark showing the Gammelmark site (marked by a black closed triangle). (B): Simplified overview of the Gammelmark section at Brunsbjerg, showing sedimentary units and OSL sampling locations (black circles). Also presented are revised quartz and K-feldspar ages (Buylaert et al., 2011). Samples 981007, 981009 and 981013 correspond to log number 5, 7 and 11 respectively. This figure is taken from Buylaert et al. (2011).

ages obtained by Murray and Funder (2003); they recalculated doses using early background subtraction (Ballarini et al., 2007), improved the cosmic ray dose rate contribution, and obtained multi-grain quartz ages of 113 ± 6 ka (981007), 126 ± 5 ka (981009), and 121 ± 4 ka (981013), which suggests that multi-grain quartz may slightly underestimate the age control. Using the fading-corrected infrared stimulated signal measured at 50 °C (IR_{50}) from K-rich feldspar extracts, Buylaert et al. (2011) dated these three samples to 120 ± 4 ka, 115 ± 4 ka and 122 ± 3 ka, respectively, and concluded that the overall average K-feldspar age (119 ± 6 ka, n = 25) for the Eemian layer was in agreement with the expected age range and not significantly different from the revised quartz result (114 ± 7 ka, n = 25). Buylaert et al. (2012) used the pIRIR(50,290) protocol (Thiel et al., 2011) on the same K-rich extracts and obtained corresponding ages of 108 ± 5 ka, 110 ± 5 ka, and 123 ± 6 ka.

The two remaining samples, H22547 and H22553, come from Sula 22 (Figure 4.2), a section located on the Sula river, Russia (see Murray et al., 2007).

They were sampled from a unit of marine sand characterized by the presence of fauna

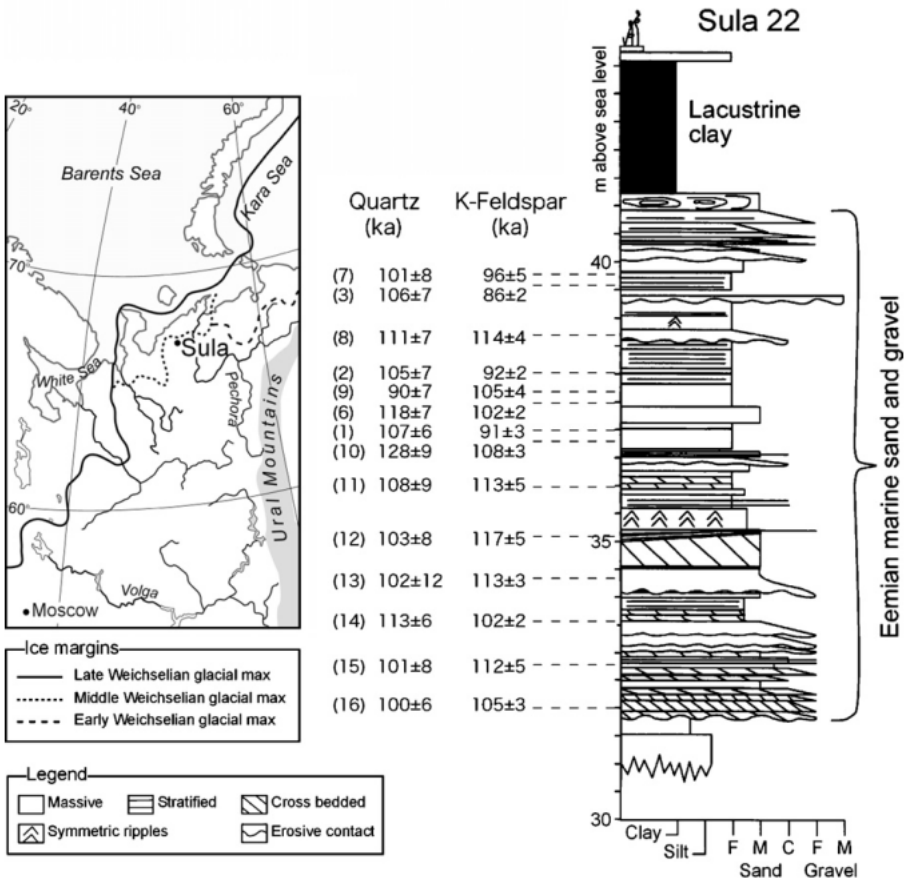

Figure 4.2: Map of a subsection of the Russian Northwest coast, showing the location of Sula (black filled circle). Also shown are the Sula 22 sedimentary units along with quartz ages from Murray et al. (2007), and K-feldspar IR_{50} ages from Buylaert et al. (2008). H22547 and H22553 correspond to log numbers 9 and 15 respectively. This figure is taken from Buylaert et al. (2008).

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 column, while also stating that the entire sequence was deposited within a span of 5,000 years. Using a double saturating exponential function ($D_{c1} \approx 45$ Gy and $D_{c1} \approx 450$ Gy) to fit the quartz DRCs, Murray et al. (2007) determined SAR ages of 90 ± 7 ka and 101 ± 8 ka for these two samples respectively, which indicate that SAR OSL ages underestimate the expected age. Using the SARA protocol (Mejdahl and Bøtter-Jensen, 1994), Murray et al. (2007) obtained quartz ages for the two samples of 127 ± 13 ka and 159 ± 13 ka, respectively. Buylaert et al. (2008) measured K-rich feldspar for the two samples using the IR_{50} signal and obtained fading corrected ages of 105 ± 4 ka and 112 ± 5 ka, respectively. Buylaert et al. (2012) measured K-rich feldspar extracted from sample H22553 using the pIRIR(50,290) signal and obtained an age

of 135 ± 7 ka.

Both sites (Gammelmark and Sula) have been determined to be early Eemian deposits by pollen analysis. The exact ages 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. For the purpose of our study, we adopt an early Eemian age of 128 ka for all samples presented here, and assign an uncertainty of 2 ka since samples were likely deposited over a few thousand years.

4.3 Experimental details

4.3.1 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., 2003b). 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).

4.3.2 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 4.1) 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 (Vanderberghe et al., 2008). We use the

Table 4.1: Water content and measured radionuclide concentrations used to derive quartz (Q) dose rates for the Gammelmark and Sula samples. Both the original dose rates "Q dose rate*" used in Murray and Funder (2003) and Murray et al. (2007) as well as the recalculated dose rates are given.

Lab code	Depth (cm)	w.c. (%)	^{238}U (Bq/kg)	^{226}Ra (Bq/kg)	^{232}Th (Bq/kg)	^{40}K (Bq/kg)	Q dose rate* (Gy/ka)	Q dose rate (Gy/ka)
981007	307	31	11 ± 4	15.1 ± 0.3	11.3 ± 0.3	405 ± 9	1.34 ± 0.04	1.45 ± 0.06
981009	398	26	16 ± 3	16.7 ± 0.6	13.8 ± 0.5	371 ± 15	1.36 ± 0.04	1.47 ± 0.07
981013	581	22	14 ± 4	17.8 ± 0.4	15.9 ± 0.3	431 ± 9	1.57 ± 0.04	1.70 ± 0.07
H22547	530	28	8 ± 4	4.3 ± 0.6	5.0 ± 0.5	330 ± 20	1.09 ± 0.06	1.03 ± 0.06
H22553	1020	26	12 ± 4	5.5 ± 0.6	5.9 ± 0.5	210 ± 16	0.86 ± 0.05	0.74 ± 0.04

saturated water content for dose rate calculations as in the original publications. In the case of Gammelmark, Murray and Funder (2003) settled on fully saturated water contents because "*... of the rate of cliff retreat due to coastal erosion; even only a thousand years ago, our samples would have been many tens of metres inside the cliff, and well below the groundwater table. It is very unlikely that this section was ever above the water table before this. There may have been some post-depositional compression, but the majority of this probably happened as the section accumulated (especially the initial de-watering of the Anodonta clay and sand layer, which was covered by about 7 m of sand by 125 ka). By the time the ice arrived, the sediment would have been frozen, and so incompressible. There was no detectable sign of permafrost disturbance of the faintly laminated overlying sands probably older than 116 ka, and certainly older than about 70 ka*" (Murray and Funder, 2003). In the case of Sula, Murray et al. (2007) assume that both H22547 and H22553 samples were saturated with either water or ice for most of the burial period, given that Pleistocene permafrost had degraded only recently from the site (∼8 ka, Tveranger et al., 1995) and given that permafrost could still be observed near the site.

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 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 Gy (H22553). 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.

4.3.3 Measurement protocols and dose determination

For both multi-grain and single-grain measurements, we employed a Single-Aliquot Regenerative (SAR) protocol (Murray and Wintle, 2000) including at least one recycling, one IR depletion 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. For each multi-grain dose and test-dose measurement, we stimulated aliquots for 40 s at 125 ℃ 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 (Ballarini et al., 2007). With single-grains, we stimulated each grain for 1 s at 125 ℃ 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. Individual dose estimates were derived using "Analyst v4.56" (Duller, 2015), and processed in the statistical computing environment "R" (R Core Team, 2023) using R-packages "Luminescence" (Kreutzer et al., 2012, 2022) and "BayLum" (Christophe et al., 2023). All visuals presented in this chapter were created using R-packages "ggplot2" (Wickham, 2016) and "egg" (Auguie, 2019).

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. 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 measure-

ments respectively (Thomsen et al., 2005). 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).

4.3.4 Rejection criteria and dose models

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 % ("$s_{T_n} < 20\%$")

(ii) include all grains where $s_{T_n} < 20\%$, 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):

   (a) "$s_{T_n} < 20\%$"

   (b) the recycling ratio is within 2σ of unity

   (c) the IR depletion ratio is within 2σ of unity

   (d) the IR depletion ratio is within 2σ of unity, and

   (e) the recuperation is less than 5% of the natural signal

(iii) include all grains which satisfy "$s_{T_n} < 20\%$" in combination with the D_c 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 Figure 4.3H. 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) ADM with an intrinsic over-dispersion $\sigma_m = 45\%$ (Gammelmark) and $\sigma_m = 30\%$ (Sula), (ii) CAM and (iii) BayLum (assuming a log-normal dose dispersion).

4.4 Luminescence characteristics

4.4.1 OSL stimulation curves, dose response curves

Based on multi-grain measurements, our samples appear to be fast component dominated (see Figure 4.3A), 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 Figure 4.3B), 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. Representative Sula multi-grain and single-grain dose response curves are shown in Figure 4.3C and 4.3D. Also shown are recycling, IR depletion and recuperation ratios. 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 $-0.03 \pm 0.14\%$ of the natural (n = 64). For Gammelmark (see Figure 4.3E), the multi-grain average recycling ratio is 1.004 ± 0.005, the average IR depletion ratio is 0.973 ± 0.005 and the average recuperation ratio is $0.14 \pm 0.02\%$ of the natural signal (n = 94). The Gammelmark multi-grain dose response curve (figure 4.3E) required a double-saturating exponential function to fit the data adequately. The median D_c values for the double saturating exponential fits of the DRCs are 25 Gy and 185 Gy (n = 154). Figure 4.3F shows representative single-grain OSL dose response curves for Gammelmark (981009) highlighting the usual considerable between-grain variability for single-grain measurements (e.g., Duller, 2008). Figure 4.3H shows the D_c-distribution at the single-grain level for 981009 (predicted dose of 188 Gy). 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. We consider both Gammelmark and Sula quartz suitable for OSL dating.

4.4.2 Recovery of laboratory given doses

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 was given and measured. No consistent difference could be seen between the different modes of bleaching and thus dose recovery results have been combined. Thermal transfer was assessed by measuring the dose retained in a subset of aliquots after blue LED bleaching. This dose was in all cases less than 0.7 Gy, and we conclude that thermal transfer is not significant for these 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 Figure 4.4.

Our multi-grain SAR recovered doses given to Gammelmark samples over the full range of tested given doses (25-300 Gy), with an overall mean 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 ∼14% (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 Figure 4.8) and thus the unacceptable dose recovery ratio of the Sula samples is not caused by first cycle sensitivity changes.

In the single grain dose recovery experiments, the given dose range between 70 and 250 Gy (981009 and H22553). The results are shown in Figure 4.5 and in Table 4.2.

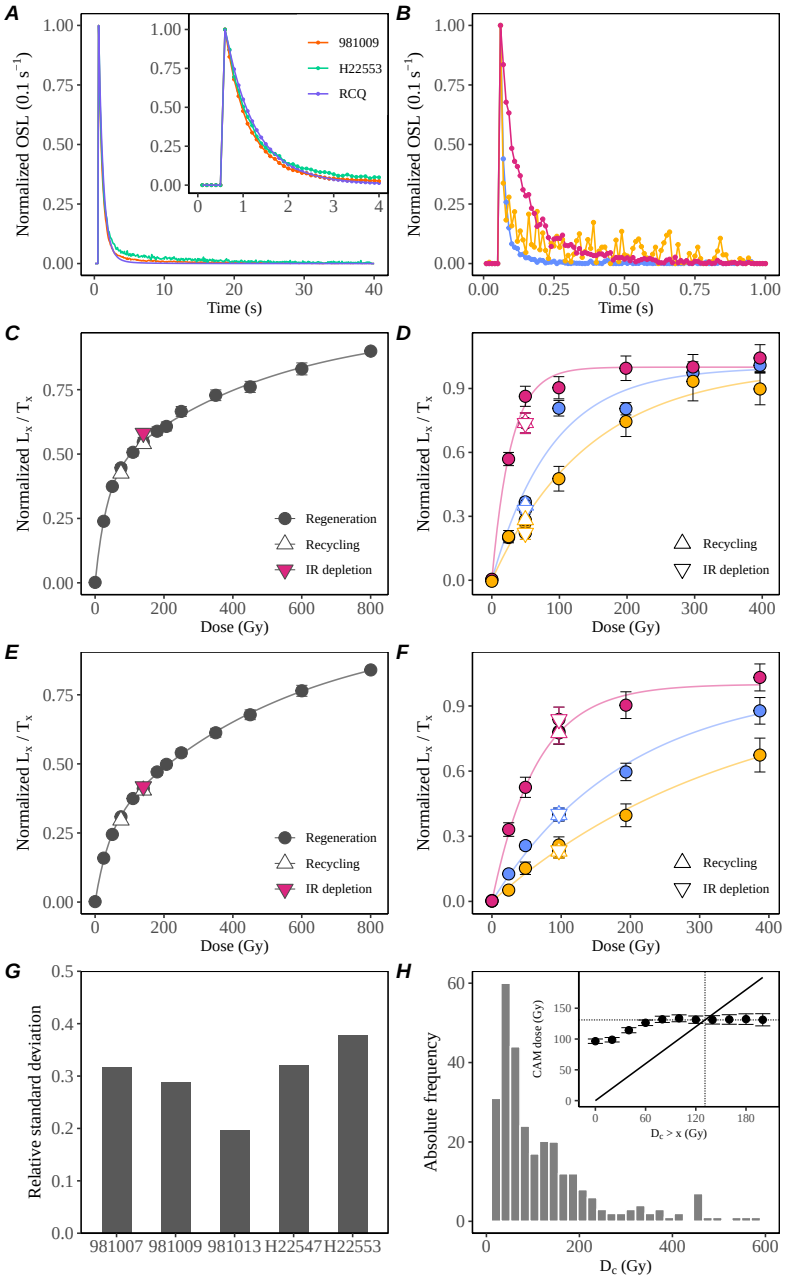

Figure 4.3: Luminescence characteristics for the Gammelmark and Sula samples. Left-hand plots all show multi-grain (MG) characteristics and right-hand plots all show single-grain (SG) characteristics. (A) Representative normalised MG stimulation curves for 981009, H22553 and Risø Calibration Quartz (RCQ). (B) SG stimulation curves for three individual grains of sample 981009. (C) Representative MG DRC from Sula (sample H22553). (D) DRCs of three individual Sula grains (sample H22553). (E) Representative MG DRC from Gammelmark (sample 981009). (F) DRCs of three individual Gammelmark grains (sample 981009). (G) MG relative standard deviation 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 and ADM doses.

Table 4.2: Overview of single-grain dose recovery results for CAM and BayLum dose models. "N" is the total number of measured grains and "n" is the number of grains giving bonded dose estimates. "n_{sat}" is the number of light-giving grains which did not give bounded dose estimates because of saturation effects. "BayLum (no sat.)" does not include grains giving unbounded dose estimates (saturated grains). Uncertainties are given at 1σ for each of our three rejection criteria schemes. Also included are over-dispersion values in % and the relative number of grains in saturation (in %).

Lab code	Given dose (Gy)	N	Rejection criteria	n	n/N	n in sat	OD (%)	CAM DR	BayLum (no sat.) DR	BayLum DR
H22553	50	2,500	$s_{T_n} < 20\%$	178	0.07	47 (21%)	30 ± 3	1.28 ± 0.04	1.38 [1.32,1.42]	1.52 [1.44,1.54]
			Std.	116	0.05	22 (16%)	25 ± 4	1.22 ± 0.04	1.28 [1.24,1.34]	1.40 [1.32,1.44]
			D_c	152	0.06	6 (4%)	30 ± 3	1.26 ± 0.04	1.32 [1.28,1.36]	1.36 [1.32,1.42]
981009	70	4,800	$s_{T_n} < 20\%$	278	0.06	124 (31%)	25 ± 2	1.08 ± 0.02	1.17 [1.15,1.19]	1.25 [1.23,1.27]
			Std.	201	0.04	95 (32%)	25 ± 2	1.08 ± 0.03	1.18 [1.15,1.20]	1.25 [1.23,1.28]
			D_c	144	0.03	5 (3%)	18 ± 2	1.21 ± 0.03	1.24 [1.22,1.27]	1.26 [1.24,1.29]
981009	100	3,400	$s_{T_n} < 20\%$	129	0.04	71 (36%)	35 ± 4	0.97 ± 0.04	1.07 [1.03,1.11]	1.19 [1.17,1.24]
			Std.	95	0.03	65 (41%)	37 ± 4	0.97 ± 0.05	1.08 [1.02,1.12]	1.21 [1.19,1.25]
			D_c	65	0.02	2 (3%)	28 ± 4	1.14 ± 0.05	1.16 [1.10,1.20]	1.16 [1.10,1.20]
H22553	110	1,800	$s_{T_n} < 20\%$	93	0.05	32 (26%)	29 ± 2	0.988 ± 0.010	1.06 [1.02,1.10]	1.16 [1.10,1.22]
			Std.	55	0.03	23 (29%)	30 ± 5	1.04 ± 0.06	1.13 [1.08,1.18]	1.21 [1.16,1.25]
			D_c	68	0.04	1 (1%)	23 ± 4	1.07 ± 0.04	1.11 [1.07,1.16]	1.14 [1.09,1.18]
981009	128	4,200	$s_{T_n} < 20\%$	156	0.04	120 (43%)	31 ± 3	0.91 ± 0.03	1.00 [0.97,1.03]	1.10 [1.09,1.13]
			Std.	131	0.03	106 (45%)	31 ± 3	0.90 ± 0.03	0.98 [0.95,1.02]	1.12 [1.09,1.16]
			D_c	76	0.02	0 (0%)	19 ± 3	1.05 ± 0.03	1.08 [1.04,1.11]	1.08 [1.04,1.11]
981009	180	4,800	$s_{T_n} < 20\%$	125	0.03	123 (50%)	46 ± 4	0.78 ± 0.04	0.96 [0.91,1.00]	1.06 [1.00,1.08]
			Std.	70	0.01	87 (55%)	52 ± 6	0.71 ± 0.06	0.96 [0.90,0.98]	1.04 [0.98,1.09]
			D_c	62	0.01	1 (2%)	27 ± 4	0.95 ± 0.04	1.01 [0.96,1.03]	1.01 [0.97,1.04]
981009	250	3,000	$s_{T_n} < 20\%$	106	0.04	198 (65%)	51 ± 5	0.68 ± 0.04	0.92 [0.88,0.93]	1.10 [1.06,1.15]
			Std.	77	0.03	163 (68%)	55 ± 6	0.65 ± 0.05	0.89 [0.82,0.92]	1.12 [1.09,1.18]
			D_c	17	0.01	1 (6%)	25 ± 8	1.13 ± 0.10	1.08 [0.97,0.74]	1.08 [1.07,1.20]

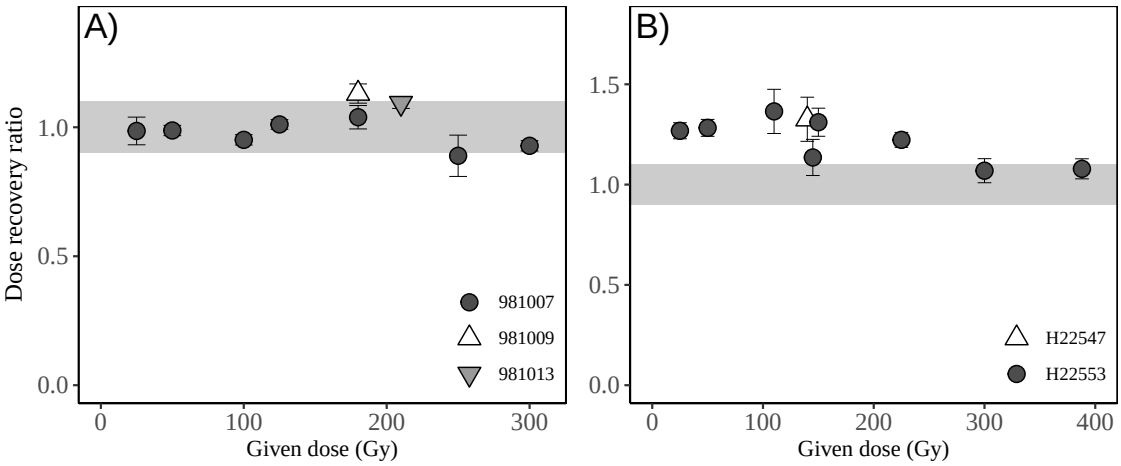

Figure 4.4: (A) Multi-grain quartz OSL SAR dose recovery experiment results for the three Gammelmark samples. (B) Multi-grain quartz OSL SAR dose recovery experiment results for the two Sula samples. Prior to giving the dose to be recovered, aliquots were bleached twice for 100 s at room temperature using either blue LEDs with a 10 ks pause between them or a daylight simulator for several hours. Grey bands indicate ±10% of unity.

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 "$s_{T_n} < 20\%$" is 0.98 ± 0.02 (n = 7) and for BayLum, the "Std." to "$s_{T_n} < 20\%$" ratio is 1.000 ± 0.013. In terms of OD, the same rejection criteria comparison yields a ratio of 1.03 ± 0.04. 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 $\text{CAM}_{sTn<20\%}$ increases from 0.98 ± 0.04 at 50 Gy to 1.66 ± 0.17 (at 250 Gy). Thus, application of the D_c criterion increases the recovered dose, but simultaneously decreases the observed OD. The ratio of OD_{Dc} to $\text{OD}_{sTn<20\%}$ decreases continuously from 1.00 ± 0.14 (at 50 Gy) to 0.49 ± 0.16 (at 250 Gy). Application of the D_c criterion also reduces the number of saturated grains to <6%. Interestingly, the dose estimated by BayLum is insensitive to the application of the D_c criterion (ratio of 0.974 ± 0.014, n = 7). 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 30 ± 3% (at 50 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. The CAM

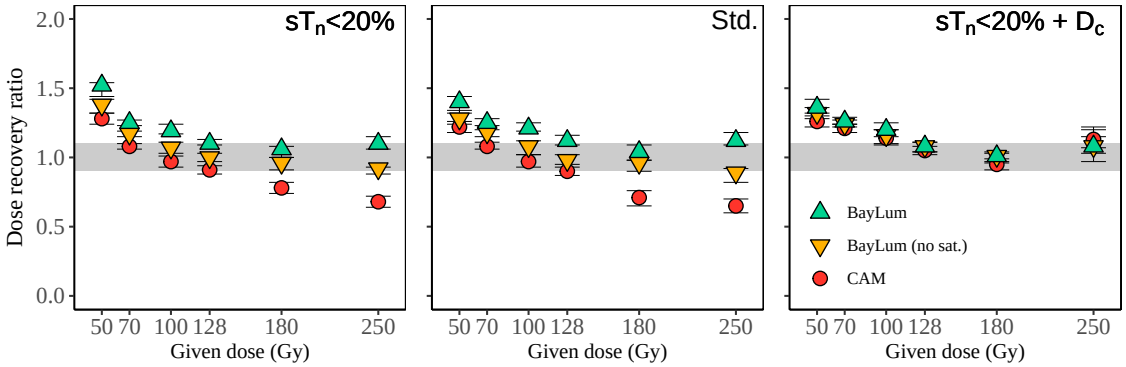

Figure 4.5: Dose recovery experiment results for samples 981009 and H22553 using our single-grain OSL SAR protocol. Samples were bleached twice for 100 s using a blue LED at room temperature, with a 10 ks pause in between. (Left): Dose recovery ratio vs Given dose (Gy) using only the $s_{T_n} < 20\%$ criterion. (Middle): Dose recovery ratio vs Given dose (Gy) using the "standard" set of criteria. (Right): Dose recovery ratio vs Given dose (Gy) using the $s_{T_n} < 20\%$ criterion in conjunction with the D_c-criterion.

dose recovery ratio decreases with given dose. At a given dose of 50 Gy, the ratio is 1.28 ± 0.04 (n = 178) and it decreases to 0.68 ± 0.04 (n = 106). Thus, 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 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. 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 $s_{T_n} < 20\%$ in BayLum, we successfully recover the given dose in the interval 110-250 Gy (average dose recovery ratio of 1.11 ± 0.02, n = 4). 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 1.00 ± 0.03, n = 5). Thus, in these dose recovery experiments, it appears that BayLum successfully recovers the given dose in the 110-250 dose range regardless of whether saturated grains are included or not. This is surprising but testifies to the power of the Bayesian approach to dose estimation.

Multi-grain quartz TT-OSL

It may be that our dating excercise approaches, or even crosses, the upper limit of dating for these quartz using standard multi-grain OSL. For this reason, we also

tested a thermally transferred OSL (TT-OSL)-protocol on our samples. TT-OSL purports to extend the age range of conventional OSL quartz dating (Wang et al., 2006).

Our TT-OSL protocol is based on a protocol presented in Arnold et al. (2015). We started with the basic framework of SAR, in which we make a series of sensitivity-corrected measurements in which each aliquot is given a number of known radiation doses in order to construct a DRC. For each regenerative cycle, we administered a known dose of radiation (the following is repeated for measurements of both the regenerative dose and the test dose) after which we heated aliquots to 260 ℃ for 10 s. Next, we stimulated aliquots using IR-LEDs to eliminate feldspar-components. We then stimulated aliquots using blue light for 100 s, which we followed up by heating samples again to 260 ℃ for 10 s after which we performed the TT-OSL stimulation using blue light at 125 ℃ for 100 s. At the end of each cycle, we hot bleached aliquots using blue light at 290 ℃ for 400 s. We used a test dose of 100 Gy.

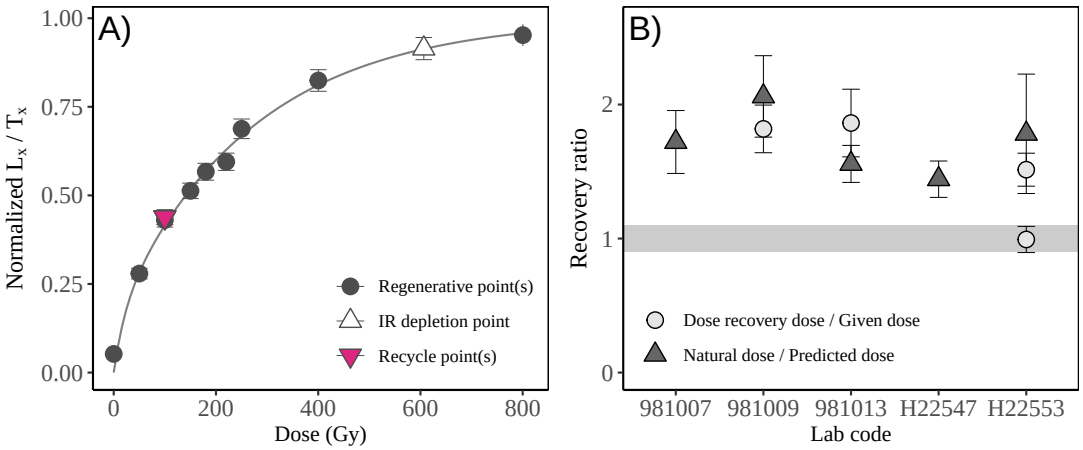

Figure 4.6: (A): Thermally Transferred OSL (TT-OSL) DRC for a representative aliquot from a dose recovery test of sample 981013. We show a double saturating exponential fit (B): (Light grey circles): TT-OSL dose recovery results for all samples using the respective predicted doses as the given dose. Labcode H22553 additionally shows a point between 0.90 and 1.10 where we gave a dose of 388 Gy. (Black triangles): the ratio between the natural doses measured and the predicted dose for each sample.

We show in figure 4.6A that the TT-MG OSL DRC grow to at least 800 Gy, and that a double saturating exponential fit is suitable in this case. Unfortunately, it is clear from figure 4.6B that our TT-OSL protocol performs poorly at dose levels we expect our samples to contain. The mean dose recovery ratio is in this case 1.73 ± 0.11 (n =

Table 4.3: Summary of multi-grain quartz natural dose results. "n_{sat}" is the number of aliquots which could not give a bounded dose estimate. "n" is the number of aliquots used in equivalent dose estimation. We show results for the arithmetic average ("$Av.$"), median and BayLum as a ratio to each sample's predicted dose (see section 4.2 for further details). For BayLum we use the "lognormal_A" dose dispersion model. We also show the median D_c-value for each of the two D_c-values for the double saturating exponential fits. Individual uncertainties are the propagated uncertainties from the measured and predicted doses.

Sample	n_{sat}	n	Av.	Median	BayLum	D_{c1} (Gy)	D_{c2} (Gy)
981007	6 (11%)	32	0.92 ± 0.07	0.91 ± 0.08	0.99 [0.93, 1.06]	22	187
981009	5 (8%)	33	0.98 ± 0.07	0.98 ± 0.08	1.00 [0.91, 1.09]	28	190
981013	3 (9%)	29	0.84 ± 0.05	0.84 ± 0.05	0.91 [0.84, 0.97]	41	161
H22547	3 (10%)	26	1.15 ± 0.10	1.07 ± 0.10	1.21 [1.09, 1.34]	25	162
H22553	1 (3%)	34	1.12 ± 0.10	1.12 ± 0.10	1.21 [1.11, 1.31]	24	300

3) - indistinguishable from the ratio of the measured natural doses to the predicted dose (1.71 ± 0.11, n = 4). Arnold et al. (2015) reviews the performance of TT-OSL on 82 samples total (from 14 different studies) and reports that 84% of ages overlap with independent (or semi-independent) age controls within 2σ. But what they also report is that samples deviating from age control are all younger than 200 ka, which prompted authors to speculate that TT-OSL may be more reliable in the range extended by TT-OSL. We here make a similar observation, since our one acceptable dose recovery result - that of sample H22553, came only once we gave a dose of 388 Gy (n = 8) - much higher than our predicted doses around the 100-200 Gy. In any case, we do not discuss TT-OSL results further.

4.5 Natural equivalent doses

Natural dose distributions were measured using both large (8 mm) multi-grain (MG) aliquots (containing between 1,200 and 4,000 grains; Duller, 2008) and single grains (SG).

4.5.1 Multi-grain equivalent doses

Multi-grain doses

Table 4.3 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 % (Figure 4.7) and appear positively skewed.

This is often interpreted as a sign of significant incomplete bleaching (e.g., Mel-

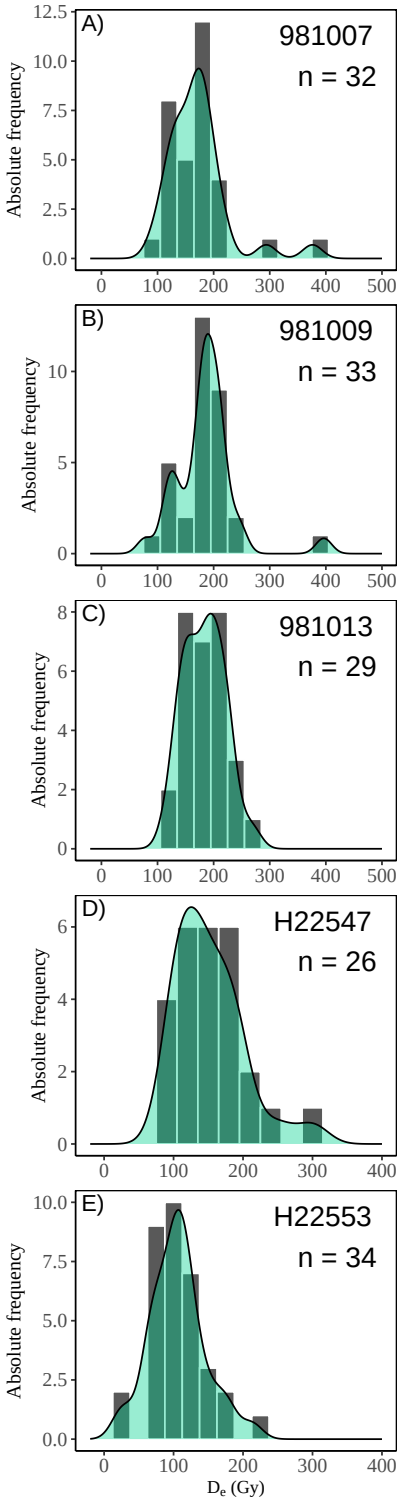

Figure 4.7: Multi-grain dose distributions for three Gammelmark samples (981007, 981009, 981013) and two Sula samples (H22547, H22553). A)-E) Simple frequency histograms with a Gaussian kernel density overlay.

lett 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 4.3).

In the single aliquot regeneration added dose (SARA) technique, aliquots containing their natural doses are given increasing additional beta doses in order to determine the relationship between the measured dose and the added dose (Mejdahl and Bøtter-Jensen, 1994). Hopefully, this relationship is suitably modelled by a linear fit whose x-axis intercept corresponds to the SARA equivalent 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. For our samples, we measured each dose using SAR. If sensitivity changes in the first cycle are negligible, the linear fit intercepts the x-axis such that the equivalent dose match exactly the dose measured when adding no dose (the slope of the linear fit is unity). In Figure 4.8, we show the SARA results for three samples (981007, 981013 and H22553).

The slope of each linear fit is consistent with unity (average slope of 1.00 ± 0.02, n = 3) and thus none of the samples appear to suffer from significant uncorrected sensitivity change occurring in the first SAR cycle. This is particularly interesting for sample H22553 for which the standard dose recovery ratio was inconsistent with unity (i.e., 1.36 ± 0.11, n = 21, at 110 Gy). The average ratio between our resulting

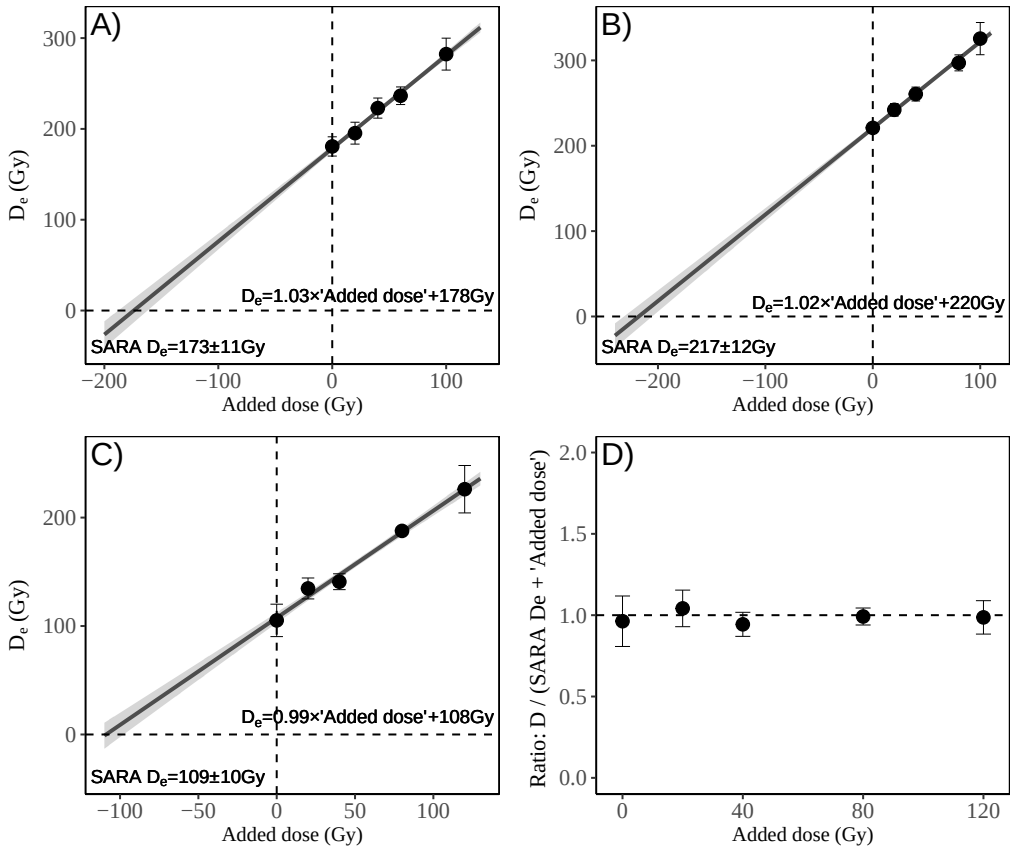

Figure 4.8: Multi-grain single aliquot regeneration added dose (SARA) results where beta doses are added on top of the natural signal and then measured using the SAR protocol. Each point is an average of at least 6 aliquots. We show a linear fit (solid line). Also shown is the 68 % confidence band A) 981007, B) 981013, C) H22553. D) Ratio of the measured dose for each natural/additive dose point to the sum of the estimated SARA D_e and the added dose for a representative sample (H22553). Direct weighting ($w_i = y_i$) was used in each regression.

SARA D_e values and the expected doses is 1.03 ± 0.05 (n = 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.

4.5.2 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., $s_{T_n} < 20\%$). However, only about 60% of these grains gave bounded dose estimates, i.e., ∼40% of the light-giving grains appeared to be in saturation. All five samples have equivalent dose distributions characterized by positive skewness and over-dispersions of more than 40% (see Figure 4.9 and Table 4.4).

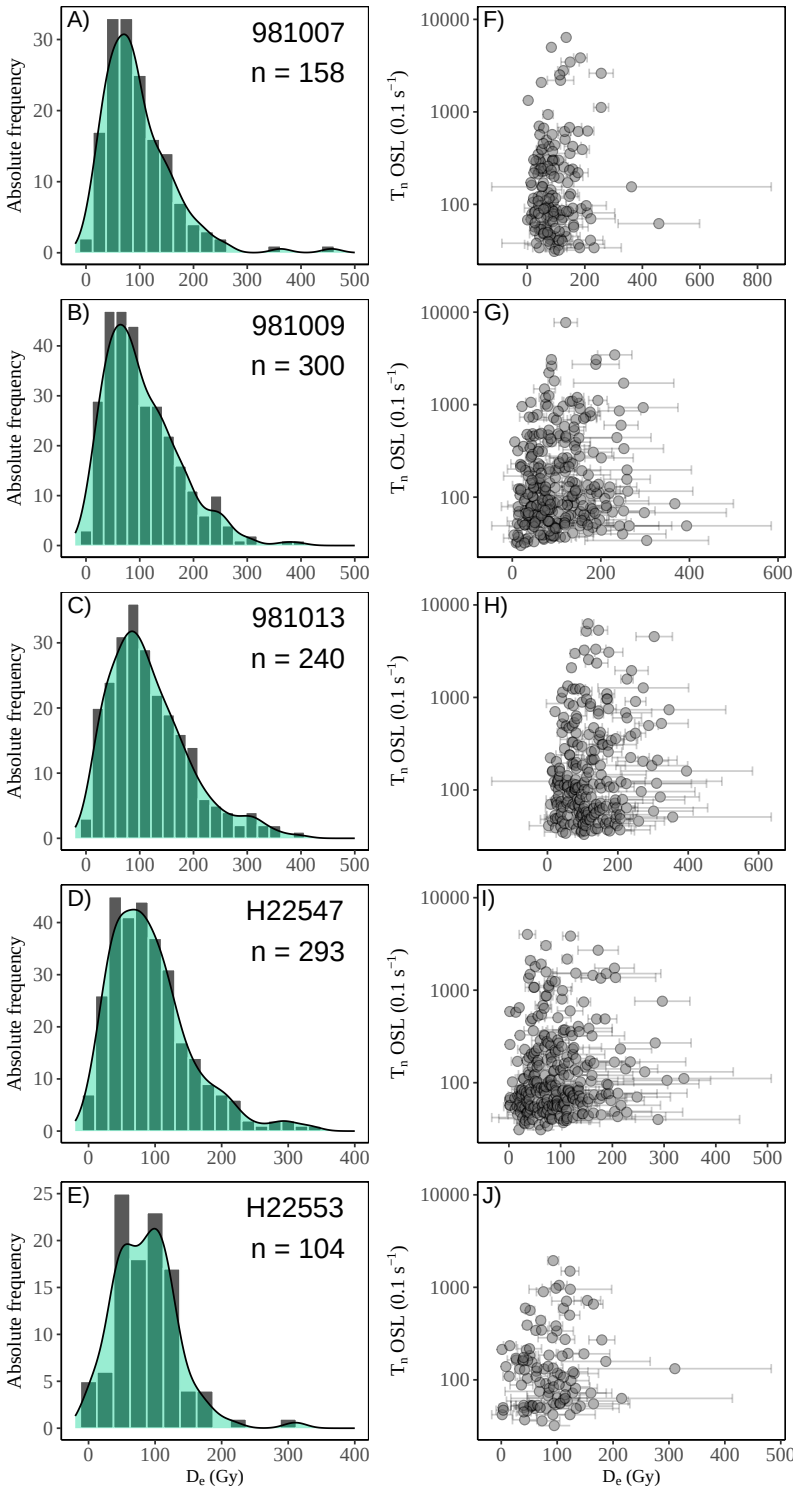

Figure 4.9: Single-grain dose distributions. A)-E) Simple frequency histograms with a Gaussian kernel density overlain. F)-J) Test dose response in the first measurement cycle (T_n) as a function of estimated dose (D_e) scatter plots. Uncertainties assigned to individual dose estimates are based on Poisson statistics, DRC fitting uncertainties and an instrument reproducibility of 2.5% per OSL measurement (Thomsen et al., 2005).

Table 4.4: Overview of single-grain results for CAM and BayLum dose models. "N" denotes the total number of grains measured while "n" represent the number of grains giving bonded 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 bonded 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).

Lab code	N	Rejection criteria	n	n/N	n in sat	OD (%)	CAM (Gy)	ADM (Gy)	BayLum$_{no\,sat}$ (Gy)	BayLum (Gy)
H22547	8,900	$s_{T_n} < 20\%$	293	0.03	169 (37%)	58 ± 3	78 ± 3	89 ± 4	105 [99,109]	149 [146,154]
		Std.	254	0.03	157 (38%)	59 ± 4	80 ± 4	91 ± 4	107 [100,112]	150 [147,154]
		D_c	158	0.02	22 (12%)	52 ± 4	93 ± 5	102 ± 5	112 [104,119]	135 [125,143]
H22553	6,500	$s_{T_n} < 20\%$	104	0.02	52 (33%)	42 ± 4	80 ± 4	83 ± 4	95 [87,101]	108 [105,112]
		Std.	75	0.01	48 (39%)	41 ± 5	86 ± 5	90 ± 5	102 [97,109]	109 [107,113]
		D_c	63	0.01	7 (10%)	32 ± 4	89 ± 5	91 ± 5	97 [90,107]	103 [99,111]
981007	7,000	$s_{T_n} < 20\%$	158	0.02	128 (45%)	64 ± 4	79 ± 5	88 ± 5	112 [104,118]	140 [128,147]
		Std.	133	0.02	122 (48%)	64 ± 5	80 ± 5	89 ± 5	116 [109,123]	138 [128,144]
		D_c	64	0.01	7 (10%)	40 ± 5	117 ± 7	117 ± 7	125 [118,132]	132 [124,140]
981009	7,600	$s_{T_n} < 20\%$	300	0.04	220 (42%)	62 ± 3	88 ± 4	96 ± 4	124 [118,129]	159 [151,165]
		Std.	252	0.03	212 (46%)	63 ± 4	88 ± 4	97 ± 4	128 [121,134]	166 [157,173]
		D_c	95	0.01	8 (8%)	46 ± 4	129 ± 7	129 ± 7	141 [132,150]	154 [142,163]
981013	4,700	$s_{T_n} < 20\%$	240	0.05	179 (43%)	62 ± 4	97 ± 5	106 ± 4	140 [133,148]	189 [176,198]
		Std.	229	0.05	171 (43%)	60 ± 4	98 ± 5	106 ± 5	141 [133,149]	191 [178,201]
		D_c	61	0.01	2 (3%)	37 ± 5	157 ± 9	157 ± 9	174 [158,183]	176 [162,187]

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 also observe the presence of a few low dose (consistent with zero) grains, the origin of which is of some mystery. 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 $s_{T_n} < 20\%$ criterion, we see from Table 4.4 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 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 4.2).

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

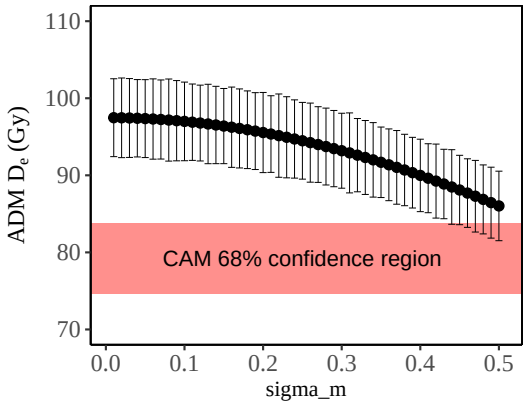

Figure 4.10: ADM equivalent dose for the data of a representative Gammelmark sample (981007) as a function of assumed relative intrinsic over-dispersion (σ_m) required for the ADM dose model for single-grain. We also show the CAM dose at the 68% confidence level for comparison.

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 $s_{T_n} < 20\%$). Similar observations have been reported numerous times in the literature (e.g., Thomsen et al., 2012; Geach et al., 2015; Guérin et al., 2015; Hansen et al., 2015; Kristensen et al., 2015; Zhao 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 $s_{T_n} < 20\%$ 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). The effect of the choice of intrinsic over-dispersion (σ_m) when estimating burial doses with the ADM in comparison with the CAM dose is illustrated in Figure 4.10 for sample 981007. The higher the value of σ_m the smaller the difference between CAM and ADM.

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 ($\text{BayLum}_{no\ sat}$) is 1.29 ± 0.05 (n = 5). Applying the D_c criterion decreases the difference between ADM and BayLum

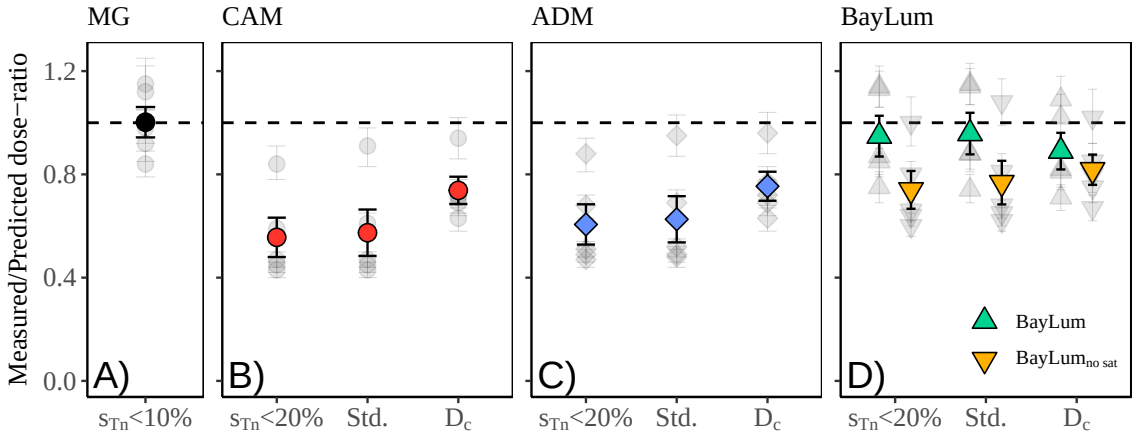

Figure 4.11: Ratios of measured equivalent doses to predicted equivalent doses as a function of rejection criteria used. A) Multi-grain quartz OSL results using the arithmetic average. B) Single-grain quartz OSL results using the Central Age Model (CAM). C) Single-grain quartz OSL results using the Average Dose Model (ADM). D) Single-grain quartz OSL results using BayLum - both with the inclusion of grains that do not give bonded dose estimates in frequentist analysis (upward facing triangles) and without these "saturated grains" (downward facing triangles). The horizontal dashed line indicates unity between the measured dose and the predicted dose (derived from the independent age control). Individual sample ratios are also shown (light grey symbols).

by ~20% (average ADM_{Dc}-to-$BayLum_{Dc}$ ratio is 0.79 ± 0.03, n = 5), partly because ADM estimates increase and partly because BayLum estimates decrease.

4.5.3 Comparisons with predicted doses from the independent age control

Figure 4.11 shows the multi-grain results obtained using the arithmetic average of the multi-grain dose results.

We observe that our multi-grain results are, on average, consistent with unity within 1σ (average MG Av.-to-Predicted dose ratio is 1.00 ± 0.04, n = 5). 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 results differ significantly from 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 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, Figure 4.11B and

4.11C clearly show that both CAM and AMD dose models severely underestimate the age control for four of the five samples with 45% and 40%, respectively, on average. Only for the sample with the lowest dose (H22553 with a predicted dose of 95 ± 6 Gy) can we recover the predicted dose within 3σ (CAM and ADM ratios of 0.84 ± 0.07 and 0.87 ± 0.07, respectively). When we apply the D_c criterion (Figure 4.11C) the average ADM_{Dc}-to-predicted dose ratio improves to 0.75 ± 0.06 (for H22553 it is 0.96 ± 0.08). However, except for H22553, this ratio is still unacceptably small. BayLum is the best performing dose model for our samples, giving an average BayLum-to-predicted dose ratio of 0.95 ± 0.08, i.e., the age control is recovered within 1σ - but only when saturated grains are included - without these grains, the ratio is 0.74 ± 0.07. However, it makes no significant difference whether we run BayLum on the D_c filtered data sets compared to the full data set, as was already demonstrated by our dose recovery results. In Figure 4.12 we plot ratios of measured equivalent doses-to-predicted equivalent 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.

4.6 Discussion

Let us now imagine a scenario in which we did not have age controls for either of the Gammelmark or Sula samples to test our findings, and our aim was to simply determine the burial age: (i) 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-

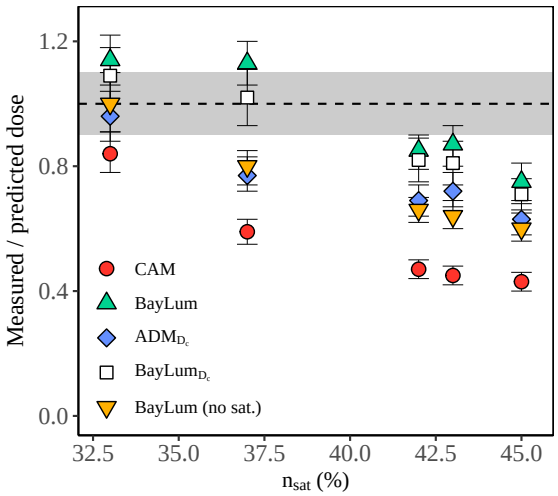

Figure 4.12: Ratios of measured equivalent doses to predicted equivalent doses as a function of the relative number of the light giving grains appearing to be in saturation (n_{sat}). For the Central Age Model (CAM, red circles) we have included all grains with $s_{T_n} < 20\%$that give bonded dose estimates. For the Average Dose Model (ADM$_{D_c}$, blue squares), we have included grains fulfilling the $s_{T_n} < 20\%$ and D_c criteria and giving bonded dose estimates. For Baylum we show the results for i) all grains with $s_{T_n} < 20\%$ (green triangles); included saturated grains, ii) Grains with $s_{T_n} < 20\%$ and fulfilling the D_c criterion (white squares) and iii) all grains with $s_{T_n} < 20\%$that give bonded dose estimates (orange triangles). Also shown is the horizontal line of recovery (solid line) and ±10 % grey band.

predicted dose ratio: 0.75 ± 0.06, n = 5). (ii) Had we instead chosen to date these samples using both conventional single-grain and multi-grain procedures, we would have been left with starkly 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 good choice for these samples. (iii) 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. But there is one single-grain methodology that does measure the expected dose accurately (on average at least), and that is BayLum. Using either only the $s_{T_n} < 20\%$ 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 (Figure 4.11. 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).

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 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).

4.7 Conclusions

In this study we investigate the accuracy of quartz OSL multi-grain and single-grain techniques on five previously published samples of known Eemian age (≈128 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 ($s_{T_n} < 20\%$) is included, (ii) only

grains with $s_{T_n} < 20\%$ 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 $s_{T_n} < 20\%$ 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. We find that single-grain analysis using CAM and standard rejection criteria severely underestimates the age controls (CAM-to-control dose-ratio of 0.57 ± 0.09, n = 5). Reducing the number of grains likely to be affected by saturation effect through application of the D_c criterion improves the accuracy, but the estimated doses still significantly underestimate the age control (ratio of 0.74 ± 0.05, n = 5). However, using BayLum, we recover the age control within 1 (mean BayLum-to-age control dose ratio of 0.96 ± 0.08, n = 5, standard rejection criteria), but only when all grains meeting the $s_{T_n} < 20\%$ criterion were analysed - including all saturated grains. On the other hand, the mean multi-grain-to-predicted dose ratio was 1.00 ± 0.04 (n = 5). We conclude that for these samples, multi-grain OSL dating of quartz is the simplest and most accurate chronometer. Our results 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].

Changelog

There are a few differences between the version of the article presented in this thesis and the one submitted to Quaternary Geochronology. These changes are primarily due to the format of the thesis, which does not have a page-limit and so does not require supplemental information (SI). In the following, I describe the Table/Figure to be moved/changed using the original article figure reference (e.g., "Fig. 2"), I then detail the action and lastly make reference to the figure within this thesis.

This is the full list of changes:

- SI1 moved, in its entirety, to section 4.2.

- Table SI1 moved to main text (Table 4.1)
- SI2 moved, in its entirety, to section 4.3.2
- Added text to reference R and used R-packages to section 4.3.3
- Fig. SI3 moved into main text (Figure 4.3)
- Fig. SI4 moved into main text (Figure 4.4)
- Table SI2 moved into main text (Table 4.2
- Fig. SI5 moved into main text (Figure 4.8)
- Fig. SI6 moved into main text (Figure 4.7)
- Fig. SI7 moved into main text (Figure 4.9)
- Added figure Figure 4.10
- Fig 2.E removed from Fig 2. and added as a separate figure (Figure 4.12)

5 Establishing an OSL chronology for Kostenki 17 - an Upper Paleolithic site by the Don River, Russia

Chapter outline

In this chapter, I derive an absolute chronology for Konstenki 17, which is an Upper Palaeolithic site located by the Don River, Russia. Using BayLum and Bayesian age-depth modelling, I present an OSL age chronology for 40 OSL samples obtained between 2017 and 2021.

This chapter will be submitted for publication in an international peer-reviewed journal as Baumgarten et al.

5.1 Introduction

Located at the centre of the East-European plain, the Kostenki-Borshchevo archaeological complex is among the largest collection of Palaeolitic sites in Europe, with ∼60 open-air sites on the west bank of the Don River, about 25 km south of Voronezh, Russia. These sites count among the most important for investigations into Anatomically Modern Humans (AMH) and the Neanderthal-to-AMH transition in an Eastern European context because of rich lithic assemblages, dynamic records of cultural evolution, well-defined stratigraphy and because similar sites elsewhere on the East European Plain are rare.

Kostenki-17 (K17) is situated on the 2nd terrace of the Don River, and it is a particularly important site for research into the Neanderthal-to-Anatomically Modern Human (AMH) transition in Eastern Europe. This is because its oldest Culture Layer (CL2) lies beneath a tephra horizon deposited by the Campanian Ignimbrite (CI) volcanic eruption (Pyle et al., 2006) which occurred 39.85 ± 0.14 ka ago (Giaccio et al., 2017). This suggests that the age of CL2 should at least be 40 ka. While there is no consensus about the techno-complex designation of the CL2 assemblages (Kozłowski, 1986; Anikovich, 1992; Sinitsyn et al., 1997; Monigal et al., 2006; Anikovich et al., 2008), it is generally believed to be of AMH origin (Dinnis et al., 2019b). Therefore, this oldest Cultural layer may represent some of the earliest AMH in the region.

Two different CL have so far been identified within K17. CL2 contains Spitsynian culture stone tools (proto-Aurignacian) and is considered to be one of the oldest Upper-Palaeolitic (UP) horizons in Europe with a developed blade industry and progressed stone age art. Recent collections include decoration elements such as pendants and beads. K17 probably records the oldest evidence of biconical drilling, found not only in arctic fox teeth and bellemnites (Palaeolithic art) but also in drilled stone tools (Stepanova et al., 2022). Another feature of CL2 is that the UP assemblage shows no presence of Middle Palaeolithic elements, meaning that the Spitsyn culture was fully developed when it arrived. The upper cultural layer (CL1) contains UP artifacts characterized by blades rather than flakes (typical of Acheulean and Mousterian industries) and a greater degree of tool standardization. CL1 and CL2 do not appear to be connected to each each other, with no evidence to suggest that lithic techniques from CL2 survived and evolved to the time of CL1.

Over the years, considerable effort has been made to radiocarbon date the cultural layers (see Table 5.1). In the late 1960s, Cherdyntsev et al. (1968) produced a radiocarbon age of between 24,606 and 23,478 years calBP (68% confidence interval using IntCal20) from humus extracted from within the upper humic bed of CL1. These authors also dated humus from the lower humic bed (within CL2) and produced a similar age of between 24,329 and 23,845 calBP (68% confidence interval, IntCal20) - an age clearly too young given that the sample was taken from below the CI tephra. Decades later, Sinitsyn et al. (1997) published additional radiocarbon ages from K17: CL2 bone material was dated to between 37,598 and 36,717 calBP, while two CL2 charcoal samples yielded ages of 39,380 - 34,451 calBP and 42,649 - 40,089 calBP (68% confidence intervals). Dinnis et al. (2019b) revisited K17 and produced five radiocarbon dates from wolf bone and fox teeth of (i) 31,185 - 30,845 calBP, (ii) 30,017 - 29,348, (iii) 37,219 - 36,200 calBP, (iv) 36,081 - 35,291 and (v) 38,551 - 36,945 calBP (68% confidence interval, IntCal20). These authors first washed the samples with acetone, methanol and chloroform solvents and then proceeded with the standard ^{14}C bone pre-treatment (i.e., ultrafiltration of bone collagen extract) protocol outlined in Brock et al. (2010). But given the chronostratigraphic positioning of CL2 below the CI tephra ($\sim$ 40 ka), all these ages appear several thousand years too young. As a result, Dinnis et al. (2019b) re-dated surplus collagen material from the initial pre-treatments, and applied the single amino acid radiocarbon dating method detailed by Deviese et al. (2018). This time, they obtained ages of 41,332 - 40,072 calBP, 41,332 - 40,072 calBP and 41,410 - 40,474 calBP (68% confidence interval, IntCal20). While Dinnis et al. (2019b) obtained credible radiocarbon ages based on the chronostratigraphic position of CL2, the authors highlight the challenge of obtaining reliable radiocarbon ages within these key layers for K17 and for the archaeological complex as a whole. Dinnis et al. (2019a) also dated a large mammal bone from CL1 to between 33,953 and 32,360 calBP (again using ultrafiltration of bone collagen extract).

It is well-documented (e.g., Devièse et al., 2021; Dinnis et al., 2019b) that radiocarbon is prone to underestimate ages older than $\sim$ 40 ka, often because of incompletely removed contamination by younger carbon. Given the importance of this site from an archaeological perspective, it is clearly desirable that an accurate absolute chronology is established.

Here, we present an Optically Stimulated Luminescence (OSL) chronology for K17. This chronology is based on 40 samples collected over a 7 m deep sediment column and is based on both multi-grain (MG) and single-grain (SG) quartz as well as K-rich feldspar measurements. To obtain burial dose estimates, we apply both standard frequentist dose models, e.g., simple arithmetic averages, the Central Age Model (CAM, Galbraith et al., 1999) and the Average Dose Model (ADM, Guérin et al., 2017) as well as a hierarchical Bayesian approach (BayLum, Combès et al., 2015; Combès and Philippe, 2017; Christophe et al., 2023). In addition, we model the age-depth relationship using three different Bayesian models: (i) *ArchaeoPhases* (Philippe and Vibet, 2020), (ii) *Bacon* (Blaauw and Christen, 2011a) and (iii) *OxCal* v4.4 (Bronk Ramsey, 2009). The accuracy of the different approaches is evaluated by comparison with the available independent age control (the CI/Y5 tephra layer).

5.2 Site description and sampling

Kostenki 17 is situated on the 2nd terrace of the Don River, about 10-15 m above the active floodplain, and down-slope from an older fluvial terrace. K17 has a complicated stratigraphy, with layers at different depths in different walls of the pit. In the following, we present a generalised description of the eastern wall of the pit, from top to bottom (see also Figure 5.1):

1. Modern soil, chernozem, 0.7-1.7 m
2. Bca Horizon of the chernozem soil, 0.1-0.4 m
3. Loess-like beige loam, 0.6-0.8 m
4. Brown loam, low humus content (ephemeral palaeosol), 0.2-0.3 m
5. Chalk pebbles, single archaeological finds (Horizon in pebbles HP), 0.25-0.5 m
6. Loess-like loam, 0.2-0.3 m
7. Brown loam, low humus content (ephemeral soil?), 0.2-0.3 m
8. Whitish loam, carbonated, dense, rare archaeological finds, 0.15-0.2 m
9. Lenses of highly humified black loam, forms the upper part of the upper humic layer (UHL), single archaeological finds, 0.02-0.15 m
10. Intensely humified loam, gleyed, with palaeosol profile, archaeological finds.

Table 5.1: Previously published ^{14}C ages for K17. Pre-treatment protocol is given where known. "AF" and "HYP" refer to "collagen ultrafiltration" (Brock et al., 2010) and "extraction of hydroxyproline from hydrolyzed bone collagen" (Brock et al., 2010; Deviese et al., 2018) pre-treatment protocols, respectively. We re-calibrated all ages using OxCal 4.4. (Bronk Ramsey, 2009) with the Intcal20 calibration curve (Reimer et al., 2020). All calibrated ages are given at 68% and 95% confidence.

Lab code	Unit	Material/Source	Pre-treatment	^{14}C uncalibrated age (yr BP)	^{14}C calibrated age		Reference
					(68%, yr BP)	(95%, yr BP)	
GIN-77	CLI	Humus		$20,000 \pm 350$	24,606 - 23,478	25,021 - 23,209	(Cherdyntsev et al., 1968)
GIN-8074	CLI	Bone		$23,000 \pm 800$	28,018 - 26,371	29,076 - 25,868	(Sinitsyn et al., 1997)
GIN-8075	CLI	Bone		$24,300 \pm 500$	28,968 - 27,917	29,784 - 27,659	(Sinitsyn et al., 1997)
GIN-8076	CLI	Bone		$21,600 \pm 600$	26,805 - 25,200	27,244 - 24,623	(Sinitsyn et al., 1997)
GrN-10511	CLI	Charcoal		$26,750 \pm 700$	31,505 - 30,200	32,915 - 29,698	(Sinitsyn et al., 1997)
- - 2756-16	CLI	Bone	AF	$28,870 \pm 380$	33,953 - 32,360	34,208 - 32,084	(Dinnis et al., 2019a)
GIN-78	CLII	Humus		$20,100 \pm 200$	24,329 - 23,845	24,710 - 23,770	(Cherdyntsev et al., 1968)
GrN-10512	CLII	Charcoal		$32,200 \pm 2,000$	39,380 - 34,851	42,210 - 33,358	(Sinitsyn et al., 1997)
GrN-12596	CLII	Charcoal		$36,780 \pm 1,700$	42,649 - 40,089	44,908 - 38,493	(Sinitsyn et al., 1997)
LE-1436	CLII	Bone		$32,780 \pm 300$	37,598 - 36,717	38,440 - 36,370	(Sinitsyn et al., 1997)
SPb-669	CLII	Bone		$28,500 \pm 300$	33,151 - 32,147	33,669 - 31,836	(Khlopachev, 2016)
OxA-30824	CLII	Tooth	AF	$26,830 \pm 250$	31,185 - 30,845	31,338 - 30,378	(Dinnis et al., 2019b)
OxA-30825	CLII	Tooth	AF	$25,480 \pm 220$	30,017 - 29,348	30,099 - 29,219	(Dinnis et al., 2019b)
OxA-32594	CLII	Bone	AF	$32,350 \pm 450$	37,219 - 36,200	38,334 - 35,730	(Dinnis et al., 2019b)
OxA-32595	CLII	Bone	AF	$31,250 \pm 400$	36,081 - 35,291	36,357 - 34,735	(Dinnis et al., 2019b)
OxA-32596	CLII	Bone	AF	$33,050 \pm 500$	38,551 - 36,945	39,209 - 36,472	(Dinnis et al., 2019b)
OxA-X-2677-56	CLII	OxA-32594	HYP	$35,650 \pm 690$	41,332 - 40,072	41,916 - 39,572	(Dinnis et al., 2019b)
OxA-X-2677-57	CLII	OxA-32596	HYP	$36,020 \pm 740$	41,766 - 40,531	42,125 - 39,817	(Dinnis et al., 2019b)
OxA-X-2717-26	CLII	OxA-30824, OxA-30825	HYP	$35,840 \pm 520$	41,410 - 40,474	41,856 - 39,952	(Dinnis et al., 2019b)

Cultural layer 1, 0.35-0.6 m

11. 10a. Lenses of highly humified black loam in the middle part of layer 10, clearly visible in the southeastern corner of the eastern wall of the 2019 excavation. Finds from cultural layer 1a, 0.02 m - 0.15 m
12. Highly humified loam, brown and black, represented by interlayered lenses with signs of cryogenic deformations. Second layer of UHL, 0.1 m - 0.5 m
13. Whitish marly loam, 0.2 m - 0.3 m
14. Thin dark lens of brown loam, 0.01 - 0.08 m
15. Brown humus loam, penetrated by small cracks filled with whitish loam, single archaeological finds, 0.15 m - 0.4 m
16. Layers and lenses of black humus loam, single archaeological finds, 0.12 - 0.42 m
17. Whitish loam with fine chalk crumbs, 0.28 m - 1.15 m
18. Brown humified loam with intermittent layers and **lenses of volcanic ash**, which in some parts forms two horizons, 0.03 m - 0.40 m
19. Loess-like loam, interlayered with lenses of fine chalk crumbs and two sustained horizons of humified brown loam (palaeosol?), 0.2 m - 0.4 m
20. Alternating layers of chalk gravel, gleyed and whitish loam. Horizon with archaeological finds in deposits under volcanic ash, 1.0 m - 1.3 m
21. Whitish loam with the inclusion of thin layers with chalk chips, 0.2 m - 0.3 m
22. Highly humified loam, dark gray brown, the lower humified layer (LHL). Upper horizon of archaeological finds of **cultural layer 2**, 0.25 m - 0.4 m
23. Brown dense loam with lenses of chalk chips, 0.3 - 0.4 m
24. Alternating lenses of dark and light gray humified loam, whitish marly loam, red sandy loam, described only in the southern and southwestern parts of the excavation pit. Lower horizon of archaeological finds from **cultural layer 2**, 0.3 m - 0.5 m.
25. Chalk pebbles. Below 6.15 m - groundwater, apparent thickness 0.4 m.

CL2a and CL2b are recognized as belonging to the same cultural layer. However, at this particular location, a thin layer of clay sits between CL2a and CL2b.

Luminescence dating samples were collected over the course of three separate field campaigns, conducted in 2018, 2020 and 2021 (see Figure 5.1). During the first two campaigns, sediment was sampled at night and placed in opaque plastic bags (∼200 - 300 g each) while a separate bag was filled for the purpose of gamma-spectrometry (∼200 g). In 2021, samples were collected in metal tubes with a diameter of 5 cm. During the field work, archaeological pits were opened and the southern and eastern walls were cleaned before sampling. The archaeological walls would have dried prior to sampling and thus to obtain a more accurate of the current water content, a separate pit was dug to sample sediment material that had not been exposed to air for any prolonged duration. This material was double bagged for return to the laboratory.

5.3 Experimental details

5.3.1 Instrumentation

We measured all samples using Risø TL/OSL DA-20 readers and single-grain quartz measurements were undertaken using Risø Single Grain Laser attachments (Bøtter-Jensen et al., 2003b, 2010). Luminescence was detected using EMI 9635QA photomultipliers. For our multi-grain quartz OSL measurements, we stimulated aliquots with blue LEDs (470 nm, ∼80 mW/cm^2) and detected the resulting OSL signals through 7.5 mm of Hoya U-340 glass filter. Samples were mounted on stainless steel discs using silicone oil covering a circular area with a diameter of ∼8 mm. For single-grain quartz measurements, we stimulated using a 10 mW Nd:YVO_4 solid-state diode laser emitting at 532 nm. Individual grains were loaded onto aluminum discs containing grain holes in a 10x10 array (Ø and depth = 300 µm), thus enabling the measurement of up to 100 grains per disc. Each disc was screened for potential contamination prior to loading. For pIRIR measurements of K-rich feldspar (KF), we loaded grains onto stainless steel cups (sample Ø = 2 mm), stimulated the aliquots with infrared LEDs (870 nm, ∼130 mW/cm^2) and detected the resulting luminescence through a 2 mm BG39 in combination with a 4 mm Corning 7-59 filter. Beta irradiations used calibrated ^{90}Sr/^{90}Y sources mounted on the readers (Bøtter-Jensen et al., 2010; Hansen et al., 2015; Autzen et al., 2022).

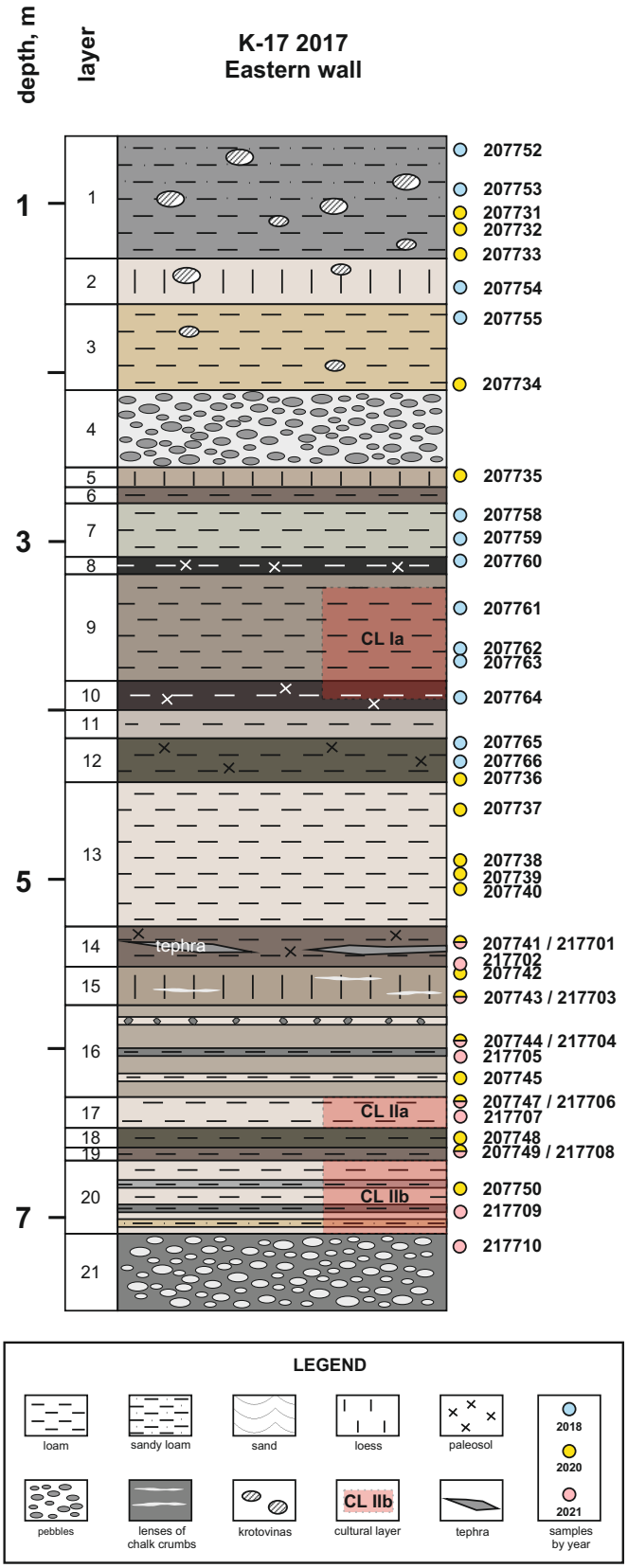

Figure 5.1: Illustration of the Eastern wall within the K17 archaeological pit.

High resolution gamma spectrometry was used to measure the bulk radionuclide activity concentrations of ^{238}U, ^{226}Ra, ^{232}Th and ^{40}K (Murray et al., 1987, 2018). Calibration was based on IAEA certified standards (^{238}U and ^{232}Th series) and high purity K_2SO_4 (^{40}K) .

5.3.2 Sample preparation

Sample material was divided into two portions: (i) material for quartz and feldspar extraction to be used for dose measurements and (ii) material for measurements of sediment dose rates.

OSL sample preparation

Extraction of quartz and K-rich feldspar were done using standard laboratory protocols (e.g., Murray et al., 2021; Mahan et al., 2023). The samples were first wet sieved to the 180-250 µm size fraction and subsequently treated with 10% HCl until any visible reaction ceased (plus one additional hour). Samples were then cleaned using water and subjected to 10% H_2O_2 for at least 24 hours at which point samples were etched in 10% HF acid for 40 minutes. We then extracted K-rich feldspar using heavy liquid separation at a density of 2.58 g/cm^3. Only very little K-rich feldspar could be extracted from the samples, and for some samples none at all. Finally, the remaining quartz-rich extract was etched in 40% hydroflouric acid for 40 minutes and wet sieved again into 180-250 µm and 90-180 µm fractions. Re-sieving was done to increase the likelihood of making true single-grain measurements.

Sample preparation for dose rate measurements

We dried dose rate samples at 50 ℃ for a minimum of one week, or until further drying resulted in no additional weight loss. The samples were then ashed at 450 ℃ for 24 hours and subsequently pulverized. The crushed samples were mixed with wax to form a cup or disc shaped sample, that was then stored for a minimum period of 21 days to establish secular equilibrium between ^{222}Rn and ^{226}Rn before measurement.

5.3.3 Measurement protocols

We employed a Single-Aliquot Regenerative-dose (SAR) procedure (Murray and Wintle, 2000) to measure both multi-grain and single-grain quartz, as well as multi-grain feldspar. For multi-grain quartz measurements (Ø=8 mm, ∼1,200 grains per aliquot), we used a preheat (before each natural or regenerative dose measurement) of 260 °C for 10 s and a cutheat of 220 °C for 0 s (before each test dose measurement).

The multi-grain quartz OSL signals were measured using blue LED stimulation for 40 s. For single-grain quartz measurements we used a preheat of 240 °C (10 s) and a cutheat of 200 °C, and measured the OSL of each grain using green laser stimulation for 1 s. A high-temperature bleach (280℃, 40 s) was performed after each step of the SAR for both multi-grain and single-grain quartz protocols. For feldspar measurements (Ø = 2 mm, ∼80 grains per aliquot) we used a $pIRIR_{225}$ measurement protocol (Thomsen et al., 2008; Buylaert et al., 2009) in which feldspar aliquots were preheated to 250 °C for 60 s and the IR signal measured at 50 °C for 100 s and then immediately after at 225 °C for 100 s. In the analysis of our multi-grain quartz OSL signals, we summed the initial 0.2 s of stimulation and subtracted a background based on the subsequent 0.4 s (early background subtraction, Ballarini et al., 2007). For single-grain quartz OSL we summed the the initial 0.06 s of stimulation and subtracted a background based on the last 0.15 s of stimulation (late background subtraction). For our feldspar IRSL signals, we summed the initial 4.4 s of stimulation and subtracted a background based on the last 10 s of stimulation.

In the SAR protocol, the natural signal (L_n) from each aliquot is measured first, followed by the OSL response (T_n) to a test dose to determine sensitivity and allow correction for any changes. An aliquot-specific dose-response curve (DRC) is then constructed by giving a series of laboratory beta regenerative doses (L_x) after each irradiation (each L_x measurement is followed by a test dose OSL measurement, T_x). In most cases, we measured four or five regenerative doses - but no less than three. To assess recuperation we also measured an additional regenerative dose at 0 Gy. For both quartz and feldspar measurements we repeated a regenerative point to determine the recycling ratio, while for quartz we repeated that same regenerative dose one additional time with an additional stimulation with IR-diodes at 50 °C for 100 s (or 125 °C for 40 s) prior to the measurement of the regenerative OSL signal to detect possible feldspar contamination (i.e., to calculate the IR depletion ratio, Duller, 2003). In order to correct for OSL sensitivity changes occurring during measurement cycles, we gave a fixed dose after each OSL measurement, i.e., a “test dose” (T_x). The chosen test dose vary between quartz samples, but ranges between 16 and 40 Gy. For feldspar measurements, the test doses ranged between 24 and 151 Gy, and was chosen to lie within 25 and 200 % of the measured natural $pIRIR_225$ dose (see Figure 5.3F). The equivalent dose (D_e) was determined by interpolation

of the sensitivity corrected natural signal (L_n/T_n) onto the laboratory constructed DRC fitted with a single saturating exponential passing through the origin, i.e.,

$$L_x/T_x = A \times (1 - e^{-D/D_c})$$

where A is the saturation value and D_c is a measure of the curvature of the DRC. Measurements were analysed using "Analyst" version 4.57 (Duller, 2015) and uncertainties assigned to individual dose estimates are based on Possion counting statistics, curve fitting errors and instrument reproducibility. For the latter, multi-grain measurements used 0.5% per OSL measurement, whereas for single-grains we used 2.5% per OSL measurement (Thomsen et al., 2005). Data were processed using the statistical software "R" (R Core Team, 2023) with R-packages "Luminescence" (Kreutzer et al., 2012, 2022) and "BayLum" (Christophe et al., 2023). All visuals were produced with R-packages "ggplot2" (Wickham, 2016) and "egg" (Auguie, 2019).

5.3.4 Dose estimation

We only derive dose estimates for individual aliquots if (i) the natural sensitivity corrected signal (L_n/T_n) are not in or above saturation of the laboratory measured DRCs, where saturation is defined to be when $L_n/T_n + s_n > A$. Here s_n is the uncertainty assigned to L_n/T_n and A is the fitted saturation value of the individual DRC, (ii) the relative uncertainty on the first (natural) test dose signal is less than 20% (i.e., $s_{T_n} < 20\%$). No additional rejection criteria were applied to the multi-grain data, but we have tested the effects of the following additional commonly used rejection criteria (here referred to as "standard rejection criteria" ("Std.")) on the single grain quartz data: (a) recycling ratios must be consistent with unity within two standard deviations, (b) the IR depletion ratio with sensitivity correction must be consistent with unity within two standard deviations and (c) the recuperation ratio must be less than 5% of L_n/T_n. We also test the effect of the D_c criterion (Thomsen et al., 2016; Singh et al., 2017) in which dose estimates originating from grains whose DRCs have D_c values less than the sample running average are not included in the burial dose estimation. We use the InterQuartile Rejection (IQR) criterion (Medialdea et al., 2014) to identify and reject individual dose values more than 1.5 times the above the upper interquartile range or 1.5 times below the lower quartile (25 percent).

For multi-grain quartz and feldspar measurements, we derive burial dose estimates using a simple arithmetic mean (see e.g., Guérin et al., 2017; Murray et al., 2021). For multi-grain quartz, we also use a hierarchical Bayesian approach (BayLum Combès et al., 2015; Combès and Philippe, 2017; Philippe et al., 2019) using a "lognormal_A" dose-dispersion model to estimate the burial dose. For single-grain measurements, we use the Central Age Model (CAM, Galbraith et al., 1999) and BayLum (using a "lognormal_A" dose-dispersion model). One of the advantages of using BayLum is that grains appearing to be in saturation using a frequentist approach can be included in BayLum analysis.

5.4 Dose rates

Infinite matrix dose rates were obtained using the conversion and grain size attenuation factors of Cresswell et al. (2018); Guérin et al. (2012), respectively. An internal quartz alpha dose rate of 0.020 ± 0.010 Gy/ka was assumed (consistent with Vanderberghe et al., 2008). For K-rich feldspar an internal alpha dose rate of 0.10 ± 0.05 Gy/ka was assumed together with an additional beta dose rate from ^{40}K and ^{87}Rb of 0.84 ± 0.04 Gy/ka. Cosmic ray dose rates are based on Prescott and Hutton (1994), using current burial depths, together with an uncertainty of 5%. The long term water content used range between 17 and 24% (see below). An uncertainty of 5% on the water content was adopted.

To derive field water contents, a pit was dug adjacent to K17, but some meters from the open face. The units of interest are all represented in this pit, although at different depths (e.g., the tephra layer is identified at around 550 cm in the main K17 pit, but at around 275 cm in the water-content pit). From these, we drive an average field water content of 19% (by weight) for all units below the Tephra (548 cm in the main K17 pit), 17% for units below the Chernozem horizons and down to, and including the tephra (548-105 cm) and lastly, 14% for the top chernozem units (92-71 cm). K17 is situated about 10-15 meters above the current floodplains of the Don. The main K17 pit extends more than 7 meters down into the ground. Consequently, several OSL samples are in close proximity, and maybe even below a seasonally fluctuating water-table. This is an effect not captured by the shallower water-content pit. Additionally, OSL samples were sampled during summer, when drier conditions prevail. Presumably, the historical water content experienced by

Table 5.2: K17 water content (w.c.) measurements from a pit dug separately to the main K17 pit for the purpose of water content measurements. Samples were transported the laboratory, weighed, dried and then re-weighed. Water content is taken to be the percent loss of weight from drying.

Sample	Depth cm	Layer	w.c. %
1	25	Chernozem	12
2	50	Chernozem	15
3	75	Loess	15
4	100	Loess	16
5	125	Paleosol	18
6	150	Silt	16
7	175	Silt with chalk	15
8	200	Silt with chalk	16
9	225	Silt with chalk	16
10	250	Chalk layer	17
11	275	Tephra layer	20
12	300	Laminated silt and clay	22
13	325	Laminated silt and clay	18
14	350	Laminated silt and clay	17
15	375	Laminated silt and clay	18
16	400	Sand (Jurassic sand)	7.2

these samples were higher, and so we chose to increase the water content figures used for dose rate calculations by ∼25% from the values obtained from the water content pit. The water contents used by each sample is shown in Table 5.3.

Figure 5.2 shows how the radionuclide concentrations, and the dry beta and gamma as well as total dose rates vary as a function of depth. For quartz the total dose rates range between 0.94 ± 0.05 Gy/ka and 2.78 ± 0.15 Gy/ka. To ensure that our dose rate measurements are reproducible, we collected 5 sample pairs, where each pair was taken at the same depth (at 538, 570, 591, 633 and 658 cm) and spatially close. Consequently we are able to compare our dose rate estimated between these duplicate samples. A priori, we would expect that each pair should give similar dose rates. This expectation is satisfied for four of the five pairs (see Table 5.3). But for the sample pair 207741/217701 (at a depth of 538 cm) we see a significant difference (1.54 ± 0.08 Gy/ka vs 1.82 ± 0.09 Gy/ka for 207741 and 217701, respectively). The ratio between the two dose rates is 1.18 ± 0.08. However, the ages calculated for these two samples are consistent with each other.

Table 5.3: Dose rate summary for the 40 sediment samples. "Current" (field) water content (w.c.) as well as the "Assumed" water content used in the calculation of total dose rates are given (see text for details). Radionuclide concentrations were measured using high resolution gamma spectrometry (Murray et al., 1987, 2018). The total dose rates for quartz (Q) and K-rich feldspar (KF) include contributions from cosmic rays, internal dose rates and the assumed water content.

Lab Code	Depth	Water content (%)		Radionuclide concentration (Bq/kg)				Total dose rates (Gy/ka)	
	(cm)	Current	Assumed	^{238}U	^{226}Ra	^{232}Th	^{40}K	Q	KF
207752	71	14	17	26 ± 7	36.2 ± 0.7	37.7 ± 0.6	583 ± 13	2.78 ± 0.15	3.77 ± 0.16
207753	92	14	17	19 ± 5	31.9 ± 0.5	31.9 ± 0.5	501 ± 10	2.42 ± 0.13	3.41 ± 0.14
207731	105	17	22	25 ± 7	26.6 ± 0.5	28.5 ± 0.6	422 ± 9	2.08 ± 0.11	3.07 ± 0.12
207732	115	17	22	24 ± 5	20.9 ± 0.4	21.3 ± 0.4	322 ± 6	1.57 ± 0.07	2.56 ± 0.10
207733	135	17	22	18 ± 2	24.8 ± 0.4	22.4 ± 0.3	345 ± 5	1.68 ± 0.08	2.68 ± 0.10
207754	150	17	22	4 ± 9	23.2 ± 0.7	23.4 ± 0.7	391 ± 14	1.79 ± 0.09	2.78 ± 0.11
207734	210	17	22	20 ± 2	25.4 ± 0.3	27.5 ± 0.3	428 ± 6	1.96 ± 0.10	2.95 ± 0.11
207735	260	17	22	22 ± 3	24.6 ± 0.6	6.7 ± 0.4	174 ± 7	1.00 ± 0.05	1.99 ± 0.08
207758	282	17	22	17 ± 2	22.4 ± 0.3	17.4 ± 0.3	300 ± 5	1.44 ± 0.07	2.43 ± 0.10
207759	299	17	22	20 ± 6	24.1 ± 0.5	20.0 ± 0.5	355 ± 8	1.64 ± 0.08	2.63 ± 0.10
207760	311	17	22	23 ± 3	31.7 ± 0.4	29.0 ± 0.4	507 ± 7	2.24 ± 0.11	3.23 ± 0.13
207761	339	17	22	18 ± 2	24.5 ± 0.4	22.0 ± 0.3	383 ± 6	1.73 ± 0.09	2.73 ± 0.11
207762	362	17	22	20 ± 8	22.5 ± 0.6	22.3 ± 0.6	377 ± 9	1.70 ± 0.09	2.69 ± 0.11
207764	390	17	22	21 ± 5	23.2 ± 1.1	18.4 ± 0.8	280 ± 12	1.40 ± 0.07	2.39 ± 0.10
207765	418	17	22	13 ± 6	20.8 ± 0.5	14.7 ± 0.5	252 ± 9	1.24 ± 0.06	2.24 ± 0.09
207766	430	17	22	21 ± 7	29.7 ± 0.7	28.7 ± 0.6	513 ± 14	2.21 ± 0.11	3.20 ± 0.13
207736	439	17	22	23 ± 7	35.9 ± 0.7	29.7 ± 0.6	461 ± 13	2.17 ± 0.11	3.16 ± 0.13
207737	458	17	22	15 ± 7	20.6 ± 0.6	15.1 ± 0.4	253 ± 9	1.25 ± 0.06	2.24 ± 0.09
207738	488	17	22	15 ± 1	20.1 ± 0.2	15.4 ± 0.1	252 ± 2	1.24 ± 0.06	2.23 ± 0.09
207739	496	17	22	29 ± 8	23.4 ± 0.6	17.9 ± 0.6	288 ± 9	1.40 ± 0.07	2.39 ± 0.10
207740	507	17	22	15 ± 5	21.0 ± 0.4	17.9 ± 0.4	270 ± 8	1.33 ± 0.07	2.32 ± 0.09
207741	538	17	22	20 ± 2	25.5 ± 0.4	22.7 ± 0.3	306 ± 5	1.54 ± 0.08	2.53 ± 0.10
217701	538	17	22	11 ± 9	34.4 ± 0.7	33.8 ± 0.7	311 ± 9	1.82 ± 0.09	2.81 ± 0.11
217702	552	19	24	21 ± 2	22.9 ± 0.3	23.6 ± 0.3	338 ± 5	1.57 ± 0.08	2.56 ± 0.10
207742	558	19	24	27 ± 6	23.3 ± 0.4	22.0 ± 0.5	351 ± 7	1.58 ± 0.08	2.57 ± 0.10
207743	570	19	24	25 ± 9	22.3 ± 0.8	17.2 ± 0.6	301 ± 12	1.38 ± 0.07	2.37 ± 0.10
217703	570	19	24	19 ± 2	23.6 ± 0.3	18.1 ± 0.2	278 ± 4	1.35 ± 0.07	2.34 ± 0.09
207744	591	19	24	14 ± 2	17.5 ± 0.3	12.7 ± 0.2	179 ± 4	0.95 ± 0.05	1.95 ± 0.08
217704	591	19	24	17 ± 2	17.8 ± 0.3	12.2 ± 0.2	176 ± 4	0.94 ± 0.05	1.94 ± 0.08
217705	611	19	24	16 ± 2	18.6 ± 0.3	14.1 ± 0.2	208 ± 4	1.06 ± 0.05	2.05 ± 0.08
207745	619	19	24	22 ± 7	20.5 ± 0.5	14.2 ± 0.6	229 ± 7	1.13 ± 0.06	2.12 ± 0.09
207747	633	19	24	20 ± 2	20.6 ± 0.3	15.8 ± 0.3	228 ± 5	1.15 ± 0.06	2.14 ± 0.09
217706	633	19	24	19 ± 4	19.3 ± 0.4	15.3 ± 0.3	239 ± 7	1.16 ± 0.06	2.15 ± 0.09
217707	642	19	24	21 ± 11	29.8 ± 0.8	37.8 ± 1.0	507 ± 14	2.26 ± 0.12	3.25 ± 0.13
207748	652	19	24	30 ± 3	31.0 ± 0.5	40.1 ± 0.5	536 ± 8	2.38 ± 0.12	3.37 ± 0.14
207749	658	19	24	30 ± 3	26.2 ± 0.7	33.7 ± 0.5	437 ± 9	1.98 ± 0.10	2.98 ± 0.12
217708	658	19	24	27 ± 2	25.1 ± 0.3	32.6 ± 0.3	487 ± 6	2.08 ± 0.10	3.07 ± 0.12
207750	680	19	24	30 ± 9	25.4 ± 0.7	31.8 ± 0.8	435 ± 11	1.94 ± 0.10	2.93 ± 0.12
217709	695	19	24	35 ± 3	30.6 ± 0.5	43.0 ± 0.4	508 ± 7	2.34 ± 0.12	3.33 ± 0.13
217710	717	19	24	28 ± 9	18.1 ± 0.7	25.2 ± 0.8	341 ± 12	1.53 ± 0.08	2.52 ± 0.10

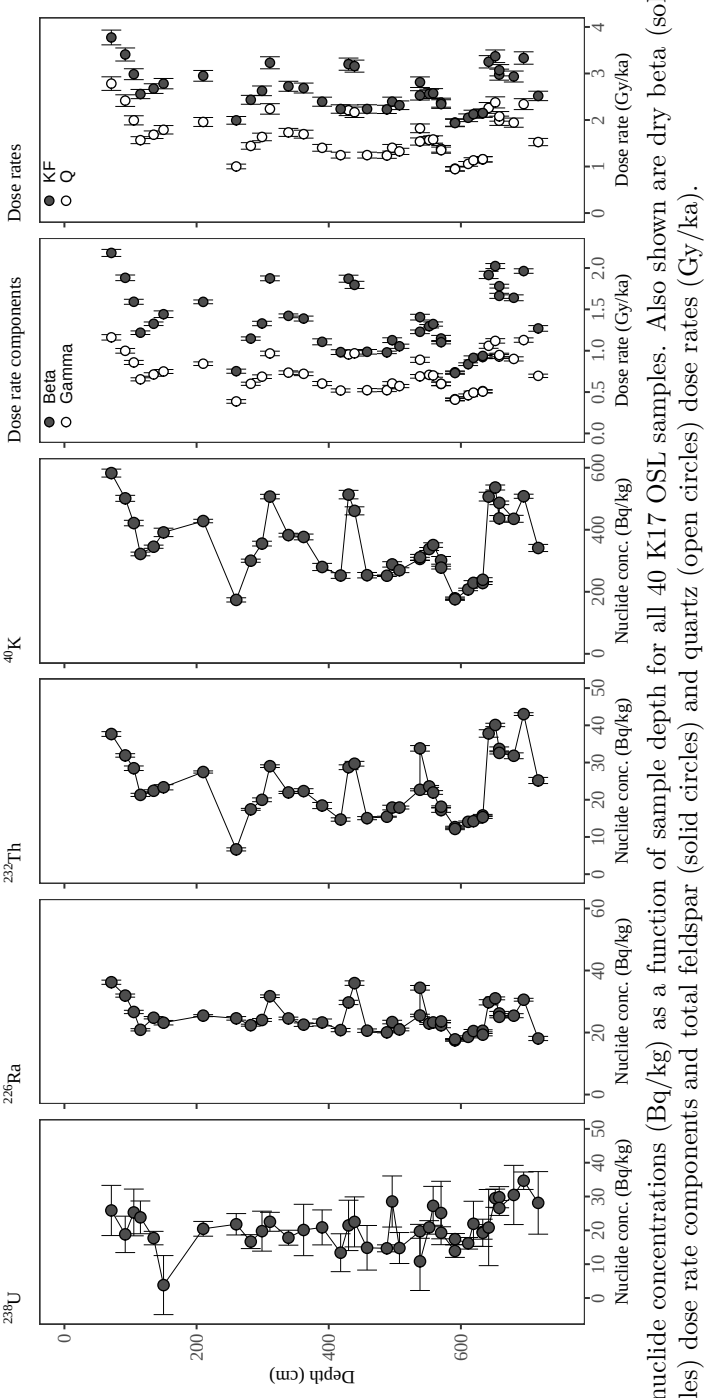

Figure 5.2: Radionuclide concentrations (Bq/kg) as a function of sample depth for all 40 K17 OSL samples. Also shown are dry beta (solid circles) and dry gamma (open circles) dose rate components and total feldspar (solid circles) and quartz (open circles) dose rates (Gy/ka).

5.5 Luminescence characteristics

In the following sections we present the luminescence characteristics for multi-grain quartz and feldspar as well as single-grain quartz.

5.5.1 Multi-grain quartz luminescence characteristics

In Figure 5.3A a representative multi-grain (MG) OSL curve from sample 207744 is shown. The MG quartz OSL signal is dominated by the fast component as can be seen from the comparison with the OSL signal obtained from a known fast component dominated sample (Risø calibration quartz, Hansen et al., 2015).

To test the applicability of our chosen SAR protocol, a series of dose recovery experiments (Murray, 1996) were carried out. Here the natural quartz OSL signal (multi-grain and single-grain) is first bleached twice using the blue LEDs for 100 s each at room temperature (with an intervening pause of 10 ks) before a known radiation dose is given to the sample. The given dose is then measured, and if the ratio between the measured dose and the given dose is at or close to unity, the chosen measurement protocol is able to recover a laboratory dose given prior to any thermal treatment adequately. Here, we regard the protocol as sufficiently accurate if the dose recovery ratio is consistent with ±10% of unity (Wintle and Murray, 2006). We tested different SAR protocols on sample 207745 using preheat temperatures ranging between 200 and 280 ℃ (the cutheat temperature was always 40 ℃ lower than the preheat temperature) with four aliquots per treatment and a given dose of 73 Gy. As can be seen from Figure 5.3C, we are able to recover the given dose over the full range of tested preheat temperatures - both using the arithmetic average and BayLum. We then further evaluated our multi-grain quartz by performing a series of dose recoveries with given doses ranging between 18 and 110 Gy, for 12 different samples. We used a SAR with a preheat/cutheat of 260/220 ℃. The results of these dose recovery experiments are presented in Figure 5.4. The dose recovery ratio over the full given dose range is 0.99 ± 0.02 and 0.98 ± 0.02 (n = 13) for the arithmetic average and BayLum, respectively.

We also measured the natural dose (sample 207764) using preheat temperatures ranging between 200 and 280 ℃ and four aliquots per preheat temperature. Ideally, the measured dose should be constant in this temperature range. If a systematic increase is observed it could indicate that a more rigorous heat treatment causes

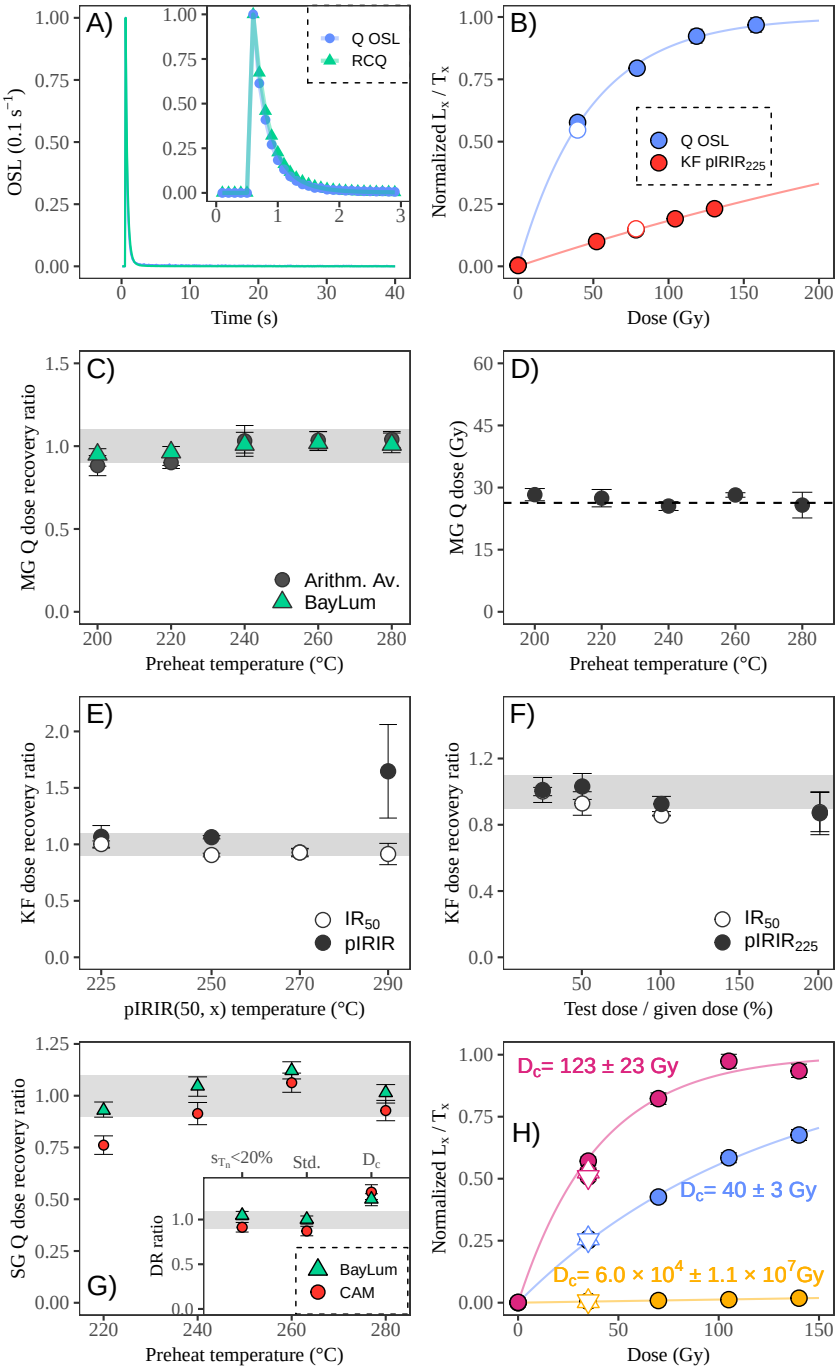

Figure 5.3: Luminescence characteristics for multi-grain (MG) and single-grain (SG) quartz (Q) and K-rich feldspar (KF). (A) Representative normalized MG Q OSL stimulation curves for sample 207744 and Risø calibration quartz (RCQ, Hansen et al., 2015). (B) Representative MG dose response curves (DRCs) for Q and KF (labcode: 207744). The DRCs have been fitted with a single saturating exponential and are normalised to the saturation value A. Recycling points are shown as open symbols. (C) MG Q dose recovery (labcode: 207745) and a given dose of 73 Gy. (D) MG Q preheat plateau (labcode: 207764). (E) KF dose recovery as a function of the 2$^{\text{nd}}$ IR stimulation temperature using a given dose of 26 Gy and a test dose of 13 Gy (labcode: 207752). The preheat temperature was 25 °C higher than the 2$^{\text{nd}}$ IR stimulation temperature. (F) KF pIRIR$_{225}$ dose recovery ratios as a function of test dose size for a given dose of 26 Gy (labcode 207752). (G) SG Q dose recovery for a given dose of 73 Gy (and a test dose of 37 Gy) as a function of preheat temperature (labcode: 207745). Dose recovery ratios are calculated using the Central Age Model (CAM Galbraith et al., 1999) and BayLum (Combès et al., 2015). The inset shows the dose recovery results for a preheat of 240 °C for three investigated rejection criteria schemes: "$s_{T_n} < 20\%$", "Std." and "D_c". (H) Representative single-grain quartz DRCs (labcode: 207745). The data have been normalized to the saturation value A. Recycling and IR depletion ratios are shown as open circles and triangles, respectively. The D_c values for the individual DRCs are given in the Figure. For all figures, grey bands indicate the range from 0.90 to 1.10. Uncertainties on individual data points are given as ± 1σ.

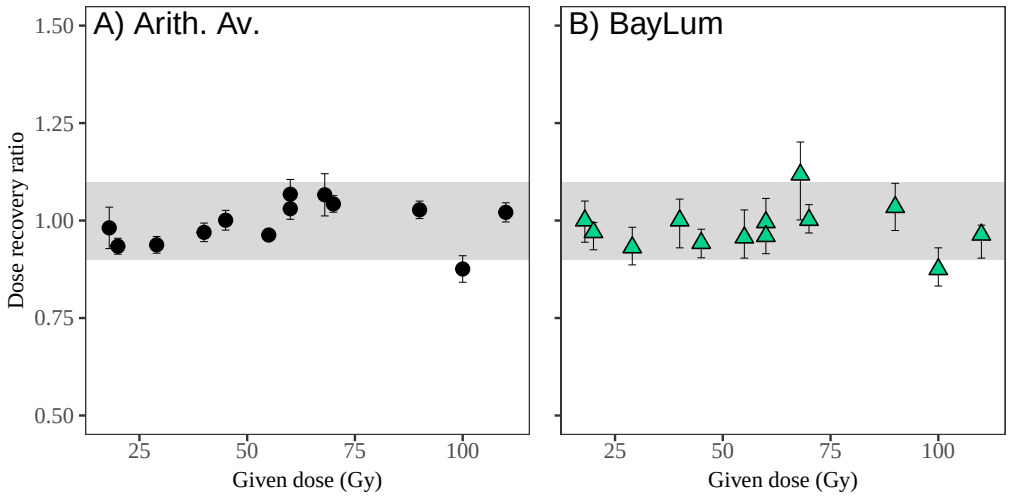

Figure 5.4: Multi-grain dose recovery results. (A): Dose recovery as a function of given dose, using the arithmetic average (Arith. Av.) to derive an estimate of the central tendency. (B): Dose recovery as a function of given dose, using BayLum to derive an estimate of the central tendency.

a transfer of charge into the OSL trap. However, this does not appear to be a dominant effect for the K17 samples, as the equivalent dose estimate is stable over a wide preheating range (see Figure 5.3D).

For the multi-grain dose recovery and preheat plateau shown in Figure 5.3C and 5.3D, the average recycling ratio was 0.97 ± 0.01 (n = 68), the average IR depletion ratio was 0.985 ± 0.004 (n = 68) and the average recuperation was 0.75 ± 0.03% (n = 68). Thus, the SAR protocol successfully corrects for any sensitivity change that may occur throughout the measurement sequence, there is no detectable feldspar contamination and no evidence of significant charge transfer between SAR cycles that could affect equivalent dose estimation.

5.5.2 Multi-grain K-feldspar luminescence characteristics

To determine an appropriate pIRIR protocol we carried out a series of dose recovery experiments with second IR stimulation temperatures of 225℃, 250℃, 270℃ and 290℃ respectively. The temperature of the first IR stimulation was kept constant at 50 ℃ and the preheat temperature was 25 ℃ higher than the second IR stimulation temperature. First, we bleached 24 aliquots in a Hönle SOL2 solar simulator at a distance of 80 cm (∼6 times more intense than full sunlight) for 48 hours. 12 of these aliquots were given a beta dose of 26 Gy and measured using the above mentioned measurement parameters and a test dose of 13 Gy (i.e., 50 % of the given dose). The remaining 12 aliquots were used to measure any residual dose still present

in the feldspar after the 48 hours of bleaching using each of the pIRIR protocols. The measured residual doses were between 1 and 3 Gy, which we subtracted from the corresponding pIRIR doses measured in the dose recovery experiment. The resulting dose recovery ratios are shown in Figure 5.3E. The $pIRIR_{225}$ protocol perform acceptably (i.e., within ±10% of unity). We also tested the dependence of the dose recovery ratio as a function of test dose size for the $pIRIR_{225}$ protocol (see Figure 5.3F) and a given dose of 26 Gy. This protocol appears to perform well using test doses ranging between 25% and 200% of the given dose.

To account for anomalous fading, we used 13 aliquots, whose natural doses had previously been determined using the $pIRIR_{225}$ protocol. Laboratory fading rates (g-values) for the $pIRIR_{225}$ and IR_{50} (measured as part of the $pIRIR_{225}$ protocol) were determined in the usual manner (Auclair et al., 2003) using doses of 31 Gy (both L_x and T_x doses). The average $pIRIR_{225}$ g_{2days} value is 0.56 ± 0.03 %/decade and the corresponding value for the IR_{50} signal is 2.79 ± 0.11 %/decade.

5.5.3 Single-grain quartz luminescence characteristics

Quartz single-grain dose recovery experiments were done by first loading the single-grain discs with the sample and the bleaching with the blue LEDs as in the multi-grain dose recovery experiments. Figure 5.3G shows the single-grain quartz dose recovery ratios calculated using both CAM and BayLum as a function of preheat temperature for sample 207745 given a dose of 73 Gy. Here all grains giving bounded dose estimates, and passing the $s_{T_n} < 20\%$ criterion, have been included in the estimation. The L_n/T_n values of 35 grains, corresponding to 34% of all light giving grains, did not give bounded dose estimates and thus cannot be included in any frequentist estimate of the dose recovery ratio, but they can be included in the Bayesian approach applied here. All dose recovery ratios are acceptable, with the exception of the one obtained using a preheat of 220 ℃ and calculated using CAM (ratio of 0.76 ± 0.04). The over-dispersion (OD) value does not depend on the preheat temperature and is on average 36 ± 9% (n = 4). In the inset to Figure 5.3G we investigate the effect of applying the various rejection criteria to the data obtained using a preheat of 240 ℃. For this sample, applying commonly used rejection criteria ("Std.", see section 5.3.4) does not significantly change the single-grain dose recovery results. In addition, the OD does not change significantly either (data not shown). Application of the D_c criterion effectively removes grains appearing to be

in saturation (0% left) but also results in a dose recovery ratio significantly larger than unity. Quartz single-grain DRC are commonly reported to vary in shape (e.g., Yoshida et al., 2000; Duller, 2008) and this also applies to the quartz from K17 as illustrated in Figure 5.3H.

Given the luminescence characteristics summarised above, it appears that quartz and K-rich feldspar extracted from K17 are well suited for OSL dating using our chosen protocols.

5.6 Natural Doses

5.6.1 Quartz multi-grain natural doses

In total, we measured natural doses for the 40 samples using 1,079 individual multi-grain quartz aliquots. We used all aliquots in BayLum modelling. Using the arithmetic average, 1,020 were used in dose calculations. Of the unused (rejected) aliquots, 19 were in saturation (2%) and 40 were rejected as their doses were 1.5 times above or below the IQR-range derived from each respective sample. The arithmetic average MG Q equivalent dose estimates range between 19.1 ± 1.1 Gy and 106 ± 3 Gy (see Table 5.4). The average D_c value is 59.4 ± 0.9 Gy (n = 455), using only aliquots with DRCs which include regenerative doses larger than 150 Gy. The average relative standard deviation of the dose distributions is 13.7 ± 0.8% (n = 40) and the skewness range between -0.55 and 0.92 (see Figure 5.5) with an average skewness of 0.32 ± 0.05 (n = 40). A skewness value less than 0.5 indicates that the distributions are generally very symmetrical.

Figure 5.5 shows the dose distributions for samples 207732, 207701 and 207705, where the two first show the most extreme skew values of the K17 MG dose distributions. Murray and Funder (2003) argued that for high dose samples, which interpolate on the more "curving" part of the DRC, a positively skewed distribution is to be expected, and in such a case the dose median would be a better estimator of the burial dose than the arithmetic average. For all samples, the mean ratio of the median dose (data not shown) to the arithmetic dose is 0.985 ± 0.003 and thus close to 1. This is not surprising given the relatively low doses measured in this study compared to the average D_c value ($\sim$40-50 Gy), as the arithmetic average and median converge to the same value as a distribution becomes more and more symmetrical. Guérin et al. (2017) were the first to argue that multi-grain ages based

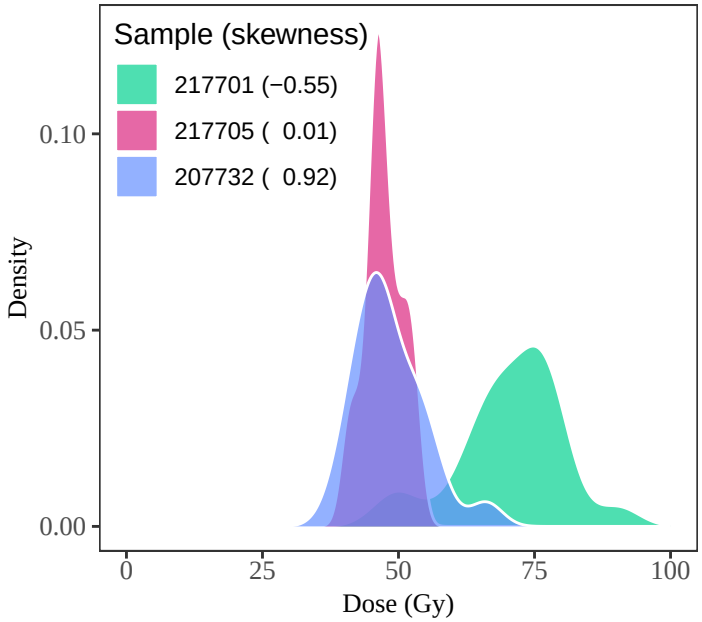

Figure 5.5: Quartz multi-grain dose distributions shown as Gaussian kernel density plots for three samples. These samples show the range of dose distribution skewness observed for K17. Sample 217701 show the largest negative skew, Sample 217705 show a skew close to zero, and sample 207732 represent the largest positive skewness.

on the arithmetic average are in general more likely to give accurate burial ages than those based on the CAM, but it is worth noting that the average ratio of the multi-grain CAM to the multi-grain arithmetic average is 0.972 ± 0.003 (n = 40) and 39 of the ratios are consistent at 95%.

All MG Q measurements have also been analysed using BayLum (see Table 5.4). The average dose ratio of BayLum to the arithmetic average is 1.028 ± 0.005 (n=40) and all individual ratios are consistent at 2σ. The average relative uncertainty is 2.7 ± 0.2 % and 3.7 ± 0.2 % for the arithmetic average and BayLum, respectively.

Thus, we do not observe any significant differences between the standard frequentist approach and the Bayesian approach for MG Q dose determination.

5.6.2 K-feldspar multi-grain natural doses

Unfortunately, it was only possible to extract K-rich feldspar for 27 of the samples and only in very small quantities; as a result only between 3 and 16 aliquots have been measured for each of the 27 samples. The $pIRIR_{225}$ average arithmetic doses range between 31 ± 5 Gy (n=15) and 330 ± 46 Gy (n=3). The corresponding numbers for the IR_{50} signal are 21 ± 3 Gy (n=14) and 130 ± 5 Gy (n=15). The relative

Table 5.4: Summary of multi-grain (MG) quartz (Q), Single-grain (SG) Q and MG K-feldspar (KF) dose results measured using the pIRIR$_{225}$ protocol. Average dose have been calculated using the arithmetic average (Av.), the Central Age Model (CAM) and BayLum (including saturated grains). All individual dose estimates pass the $s_{T_n} < 20\%$ criterion. n is the number of aliquots/grains accepted into dose calculations. "n$_{IQR}$" refers to the number of outliers rejected in multi-grain analysis using the IQR approach (see text for details). Uncertainties on individual data points are given as $\pm$ 1σ.

Labcode	Depth	MG Q						SG Q ($s_{T_n} < 20\%$)							SG Q (D_e)				pIRIR$_{225}$		IR$_{50}$	
		n$_{sat}$	n$_{IQR}$	n	Av.	CAM	BayLum	N	n_{sat}	n	n/N	CAM	OD	BayLum	n_{sat}	n	CAM	OD	n	Av.	n	Av.
	(cm)				(Gy)	(Gy)	(Gy)					(Gy)	(%)	(Gy)			(Gy)	(%)		(Gy)		(Gy)
207752	71	0	2	22	22.6 ± 1.0	22.1 ± 1.0	23.7 [22.4, 24.9]	3500	13 (5%)	254	0.07	15.6 ± 0.8	76 ± 4	23.7 [22.3, 24.9]	8 (3%)	234	16.1 ± 0.9	76 ± 4	15	31 ± 5	14	21 ± 3
207753	92	0	0	20	19.1 ± 1.1	18.5 ± 1.1	19.5 [18.2, 20.5]	3000	20 (9%)	215	0.07	16.5 ± 1	78 ± 4	27.2 [24.8, 28.8]	11 (6%)	186	17.7 ± 1.1	77 ± 5				
207731	105	0	0	24	49 ± 1.5	48.6 ± 1.4	50.6 [48.6, 52.3]															
207732	115	0	2	22	48.4 ± 1.4	47.2 ± 1.2	48.5 [46.7, 50.0]	2300	29 (15%)	169	0.07	35.2 ± 1.2	37 ± 3	42.4 [40.9, 43.7]	1 (1%)	80	42 ± 2	33 ± 4	3	76 ± 4	3	47.8 ± 0.8
207733	135	0	0	24	44.7 ± 1.3	43.9 ± 1.3	46.0 [44.0, 47.6]															
207754	150	2	0	22	56 ± 2	54 ± 2	64.5 [58.4, 67.9]	1100	22 (20%)	90	0.08	44 ± 3	48 ± 4	56.7 [52.9, 59.9]	1 (2%)	45	56 ± 5	51 ± 6	3	330 ± 46	3	128 ± 5
207734	210	0	1	22	63 ± 2	62 ± 2	64.3 [61.3, 66.5]															
207735	260	0	0	23	35.6 ± 1.0	35.2 ± 1.0	36.1 [34.8, 37.2]	2400	18 (8%)	211	0.09	28.3 ± 0.9	38 ± 3	36.2 [34.8, 37.6]	9 (6%)	135	32.2 ± 1.2	36 ± 3	10	100 ± 10	11	80 ± 18
207758	282	2	0	22	53 ± 2	52 ± 2	57.8 [54.5, 60.0]															
207759	299	1	1	32	42.3 ± 0.8	41.9 ± 0.7	43.4 [42.3, 44.3]															
207760	311	0	0	21	46.5 ± 0.3	46.4 ± 0.7	47.1 [46.0, 48.2]												3	70 ± 2	3	59 ± 2
207761	339	0	2	20	36.7 ± 0.7	36.7 ± 0.7	38.6 [37.5, 39.6]	2700	18 (9%)	187	0.07	33.4 ± 1	35 ± 3	39.0 [37.7, 40.3]	1 (1%)	142	36 ± 1.2	32 ± 3				
207762	362	0	1	23	40.9 ± 0.8	40.8 ± 0.8	42.1 [40.9, 43.0]															
207764	390	0	0	22	33.2 ± 0.5	32.9 ± 0.5	33.5 [32.7, 34.3]												5	62 ± 4	5	47 ± 6
207765	418	0	1	23	32.6 ± 0.9	32.3 ± 0.9	33.8 [32.4, 34.8]															
207766	430	0	0	10	56 ± 4	55 ± 4	58.4 [52.5, 62.5]															
207736	439	0	1	35	47.4 ± 0.7	46.9 ± 0.7	46.7 [45.7, 47.6]	2400	127 (30%)	292	0.12	41.9 ± 1	34 ± 2	53.6 [51.9, 54.9]	4 (3%)	141	42 ± 2	37 ± 3	11	82 ± 2	12	60 ± 2
207737	458	0	1	21	42.5 ± 0.9	42.2 ± 0.8	42.6 [41.5, 43.6]	2200	22 (9%)	216	0.1	33 ± 1.0	38 ± 2	39.1 [37.9, 40.3]	0 (0%)	134	36.1 ± 1.4	39 ± 3	11	94 ± 7	11	63 ± 7
207738	488	0	1	22	39.1 ± 0.7	39 ± 0.7	40.0 [39.1, 40.9]												6	81 ± 4	6	52.8 ± 1.1
207739	496	0	2	22	45.5 ± 0.8	45.5 ± 0.8	47.5 [45.8, 49.1]												5	83 ± 3	5	62 ± 3
207740	507	0	0	24	51 ± 2	50 ± 2	51.6 [49.6, 53.2]															
207741	538	1	2	21	62 ± 2	60 ± 2	66.9 [62.7, 69.9]															
217701	538	0	2	21	71 ± 2	70 ± 2	72.1 [69.6, 74.2]												4	181 ± 29	4	95 ± 6
217702	552	1	2	21	58.6 ± 1.3	58.1 ± 1.1	60.2 [58.3, 62.0]												3	99 ± 4	3	69.22 ± 0.09
207742	558	0	2	32	53.5 ± 0.9	53 ± 0.9	54.6 [53.3, 55.7]	4000	41 (14%)	249	0.06	42.1 ± 1	31 ± 2	47.3 [46.1, 48.3]	2 (1%)	153	43.3 ± 1.3	32 ± 2	7	99 ± 5	7	69 ± 2
207743	570	0	2	34	45.5 ± 0.9	45.3 ± 0.9	47.5 [46.1, 48.7]												3	96 ± 5	3	65 ± 5
217703	570	0	3	21	53 ± 2	53 ± 2	54.6 [52.7, 56.1]												4	92 ± 5	4	64.6 ± 1.2
207744	591	0	0	24	41.7 ± 1.3	41.1 ± 1.2	42.3 [41.0, 43.6]												15	81 ± 8	15	52 ± 3
217704	591	0	2	35	45.5 ± 1	44.6 ± 0.9	45.8 [44.3, 47.1]												11	97 ± 5	11	64 ± 2
217705	611	0	2	22	47.2 ± 0.7	46.8 ± 0.7	46.7 [45.4, 47.8]												14	123 ± 8	14	88 ± 6
207745	619	1	1	22	60 ± 2	59 ± 2	62.2 [59.4, 64.3]	2700	86 (26%)	247	0.09	55 ± 2	44 ± 3	72.8 [70.1, 75.0]	5 (5%)	105	62 ± 3	50 ± 4	6	81 ± 3	6	55.6 ± 1.2
207747	633	0	1	69	44 ± 0.8	43.4 ± 0.8	44.8 [43.9, 45.6]												10	91 ± 5	10	64 ± 4
217706	633	0	0	32	45.9 ± 1.3	45.7 ± 1.3	48.5 [46.4, 50.4]												3	81 ± 3	3	60 ± 2
217707	642	3	2	28	96 ± 3	96 ± 3	103.9 [99.2, 107.7]												3	162 ± 16	3	122 ± 14
207748	652	2	1	20	106 ± 2	102 ± 4	105.2 [98.9, 109.7]												6	196 ± 19	6	122 ± 11
207749	658	0	0	24	97 ± 3	90 ± 3	94.2 [89.5, 97.7]	2100	161 (32%)	344	0.16	67 ± 2	40 ± 2	97.8 [95.1, 100.3]	4 (5%)	79	83 ± 4	39 ± 4	7	200 ± 9	8	117 ± 2
217708	658	4	0	26	99 ± 3	96 ± 3	98.7 [94.2, 102.7]															
207750	680	1	1	34	80 ± 1.4	78.9 ± 1.4	81.7 [79.5, 84.0]	2800	96 (26%)	268	0.1	60 ± 2	38 ± 2	79.7 [77.5, 81.8]	2 (2%)	119	66 ± 3	38 ± 3	11	228 ± 10	11	115 ± 3
217709	695	1	1	26	105 ± 3	103 ± 3	107.4 [102.1, 111.3]												16	250 ± 16	15	130 ± 5
217710	717	0	1	32	81 ± 2	79 ± 2	82.0 [79.0, 84.6]												9	182 ± 10	9	104 ± 6

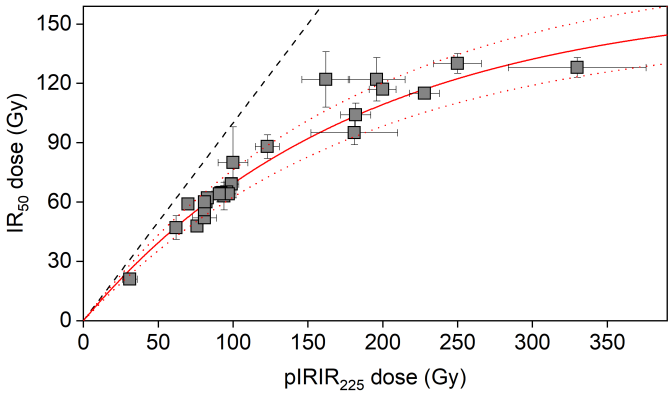

Figure 5.6: IR_{50} doses plotted as a function of $pIRIR_{225}$ doses. The solid line is a saturating exponential fitted to the data, whereas the dashed line represents the 1:1 line.

uncertainties on the individual arithmetic averages range between 2% and 16% with an average of 6.9±0.7 % (n=27) for the $pIRIR_{225}$ signal. The corresponding numbers for the IR_{50} signal are 0.0014%, 14% and 6.0 ± 1.0 % (n=27).

Figure 5.6 shows the IR_{50} doses as a function of the $pIRIR_{225}$ doses. The data have been fitted using a saturated exponential fit and does not surprisingly show a significant deviation from the 1:1 line. Both signals are prone to fading, but the $pIRIR_{225}$ (g_{2days} = 0.56 ± 0.03%/decade) signal fades much more slowly than the IR_{50} signal (g_{2days} = 2.79 ± 0.10%/decade) for these samples. On the other hand, the bleaching rate for the IR_{50} signal is known to be significantly faster than that of the $pIRIR_{225}$ signal (e.g., Thomsen et al., 2008; Buylaert et al., 2009) and thus by comparing the doses from the two signals it should be possible to determine if some of the samples could be suffering from significant incomplete bleaching. A similar approach was used by Buylaert et al. (2013), who also fitted a saturating exponential to their plot of IR_{50} against $pIRIR_{290}$ (the latter assumed not to be suffering from anomalous fading) and rejected all data below 10% of their fitted line to remove stratigraphic outliers. Applying this arbitrary but objective approach to our data, we identify three samples that may be affected by incomplete bleaching, i.e., samples 207732, -35 and -37. However, given that the functional relationship between the two signals have not been firmly established it may be a more robust approach to compare with quartz ages. This is done in section 5.7.

5.6.3 Quartz single-grain natural doses

Single-grain OSL measurements are often considered to be more accurate than multi-grain measurements, mainly because they allow detection of incomplete bleaching

and the rejection of individual grains with "aberrant" OSL characteristics (e.g., Duller, 1994; Olley et al., 1998; Jacobs et al., 2008; Armitage et al., 2011; Arnold and Roberts, 2011; Jacobs et al., 2011, 2013; Demuro et al., 2019; Liu et al., 2022; Demuro et al., 2023).

For our single-grain analysis of quartz, we measured the natural dose of 31,200 grains from 12 samples taken throughout the 7 m deep excavation pit. Of these, 3,711 grains gave detectable signals ($s_{T_n} < 20\%$), of which 767 grains (21 %) were in saturation (see Table 5.4). The natural dose distributions (including all grains with $s_{T_n} < 20\%$) are shown in Figures 5.7, 5.8 and 5.9 and most distributions appear approximately symmetrical, although a few of the younger samples appear positively skewed (particularly sample 207752 and 207753 from the Chernozem layer). There is significant scatter in the individual dose distributions, which have relative ODs ranging between 34 ± 2 % (n = 251) and 67 ± 5% (n = 215) with an average of 43 ± 3%. The single-grain dose distributions for the two Chernozem samples (207752 and 207753), collected from the top of the K17 pit, have the highest observed ODs of 64 ± 4% and 67 ± 5%, respectively, while the rest of the samples cluster around ∼40 % (average of 38.2 ± 1.4%, n = 10). For a small subset of samples, we see a notable presence of 0 Gy grains (Figure 5.8B and 5.8C, Figure 5.9A). For these three samples, a total of 11 grains have dose estimates less than 1 Gy (out of 785 unsaturated grains). A total of 14 grains have dose estimates less than 2 Gy. As expected, the relative number of saturated grains systematically increases with CAM dose (although sample 207736 has an exceptionally high number of saturated grains).

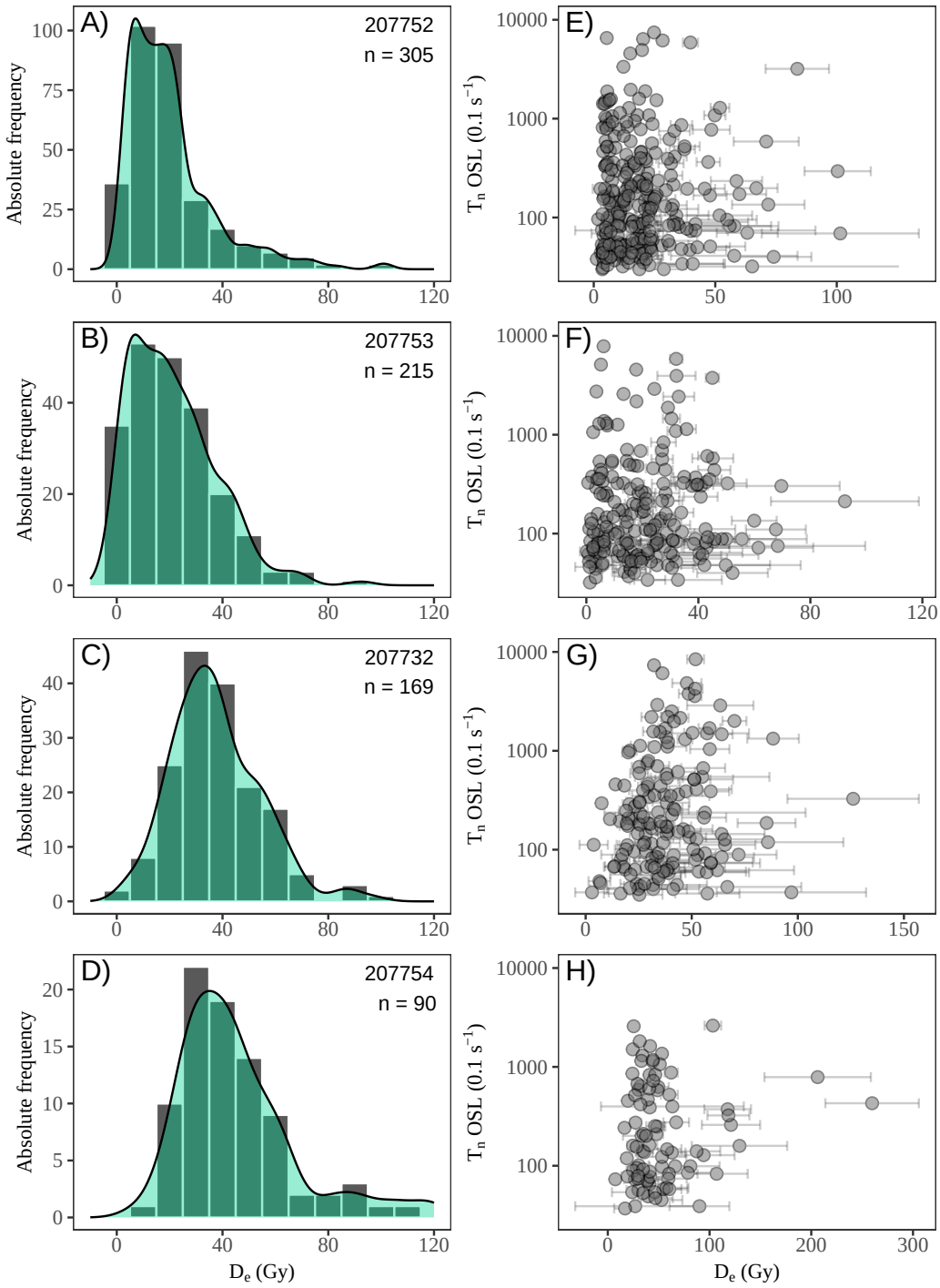

Figure 5.7: Dose distributions and natural test dose as a function of dose for samples 207752, 207753, 207732 and 207754. All included grains pass the $s_{T_n} < 20\%$ criterion. (A-D): Dose distribution histograms with Gaussian kernel density overlain. (E-H): Scatter plots of natural test dose response as a function of estimated dose. Points are semi-transparent, meaning that a high density of points on top of eachother will produce a darker shade.

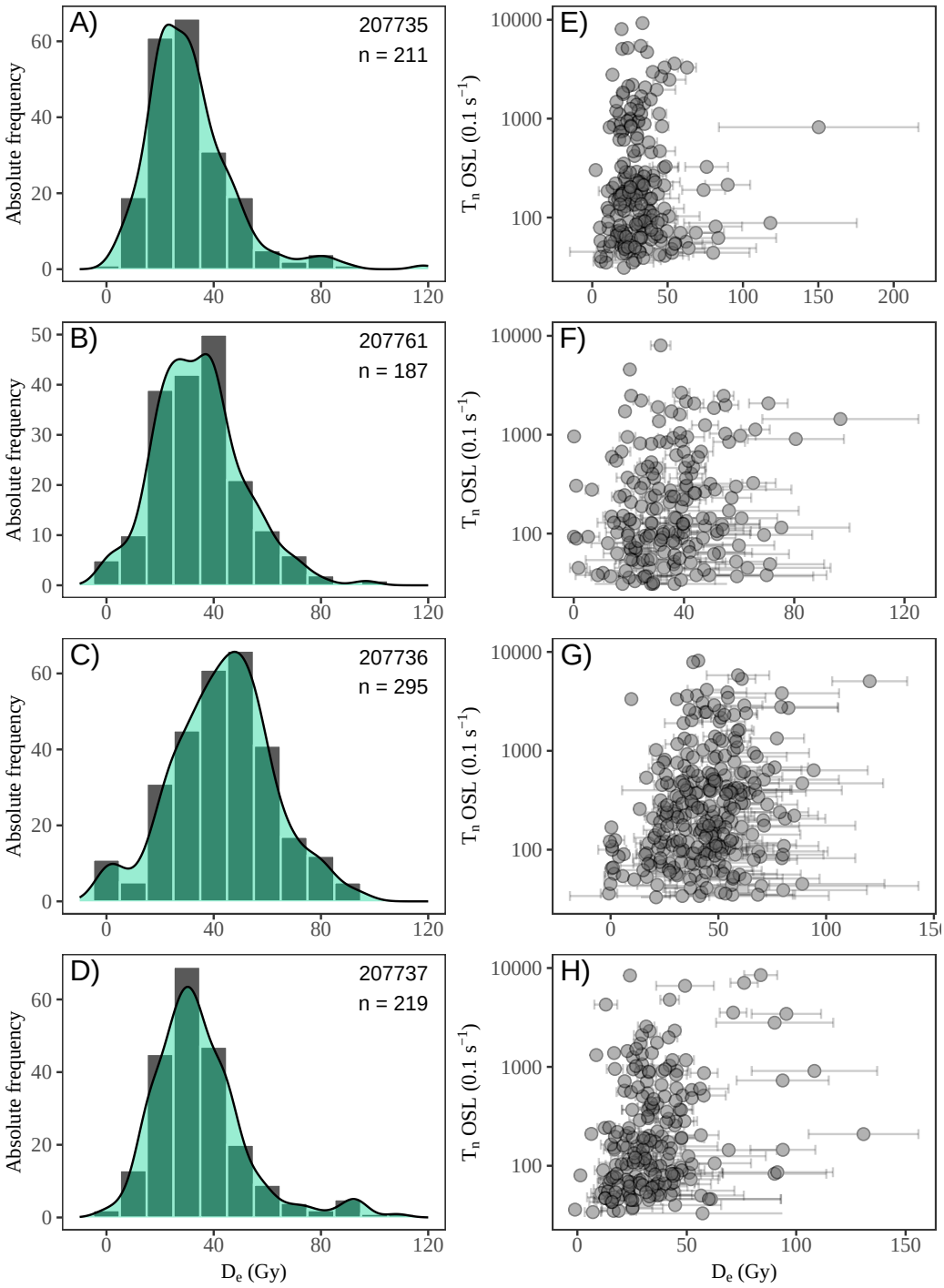

Figure 5.8: Dose distributions and natural test dose as a function of dose for samples 207735, 207761, 207736 and 207737. All included grains pass the $s_{T_n} < 20\%$ criterion. (A-D): Dose distribution histograms with Gaussian kernel density overlain. (E-H): Scatter plots of natural test dose response as a function of estimated dose. Points are semi-transparent, meaning that a high density of points on top of eachother will produce a darker shade.

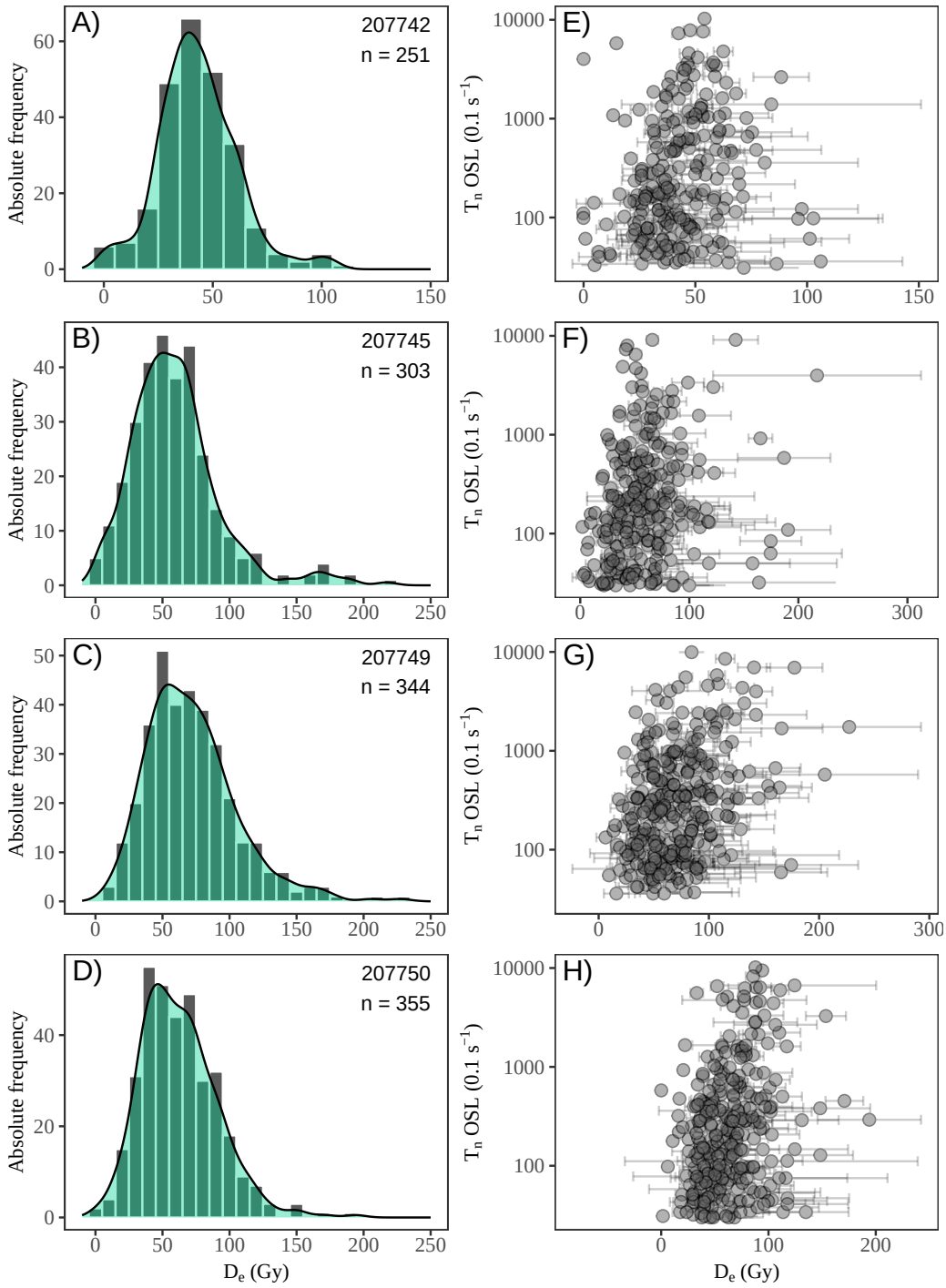

Figure 5.9: Dose distributions and natural test dose as a function of dose for samples 207742, 207745, 207749 and 207750. All included grains pass the $s_{T_n} < 20\%$ criterion. (A-D): Dose distribution histograms with Gaussian kernel density overlay. (E-H): Scatter plots of natural test dose response as a function of estimated dose. Points are semi-transparent, meaning that a high density of points on top of each other will produce a darker shade.

The effect of applying various rejection criteria schemes on the number of accepted grains (n), the CAM dose, and the relative over-dispersion (OD) compared to simply including all grains with $s_{T_n} < 20\%$ is shown in Figure 5.11. The application of

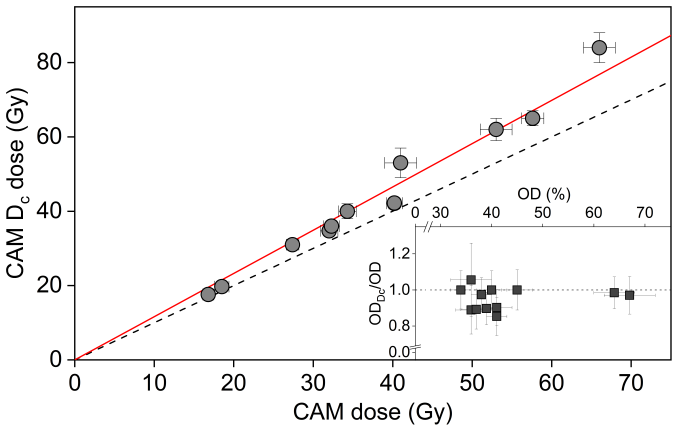

Figure 5.10: Effect on CAM dose of applying the D_c criterion on the single-grain quartz data. The solid red line shows a linear fit to the data. The dashed line indicate the 1:1 line. Inset shows the corresponding data for the relative over-dispersion (OD). The dashed line indicates unity.

popular rejection criteria, here designated "Std." (i.e., $s_{T_n} < 20\%$ recycling and IR depletion ratios both within 2σ of unity, and recuperation less than 5% of L_n/T_n), only change the estimated CAM dose (ratio of 1.041 ± 0.009, n = 12) and the OD (ratio of 0.93 ± 0.02, n = 12) slightly. The criterion that is causing this difference is the IR depletion ratio criterion. However the average number of accepted grains is reduced by $30 \pm 2\%$ and thus the relative uncertainty of the derived dose estimates is, on average, increased by 7%.

The average ratio of SG Std. CAM dose to MG quartz arithmetic dose is 0.83 ± 0.02 (n = 12). Application of the D_c criterion reduces the difference between single-grain CAM and multi-grain quartz doses, but does not completely remove it (average ratio is 0.89 ± 0.03). This is not surprising as the D_c criterion is not expected to have a significant effect on samples with a relatively low equivalent dose compared to the D_c. Figure 5.10 demonstrates this - only for CAM doses larger than ~40 Gy does the D_c criterion significantly change the estimated dose. The solid red line (slope 1.16 ± 0.03) is a linear fit to the data. The OD is not significantly affected by application of the D_c criterion (average ratio of 'with D_c criterion' to 'without D_c criterion' of 1.01 ± 0.02).

Based on the above, we now use all data for which $s_{T_n} < 20\%$. The average ratio of single-grain CAM to multi-grain quartz arithmetic average is 0.80 ± 0.02 (n = 12). The relative number of saturated grains in the twelve samples vary from 6 % to 32 % (Table 5.4).

Although CAM is widely used in the field of OSL dating, it has been pointed out (e.g., Guérin et al., 2017) that this dose model estimates the median (or geometric mean) of a lognormal distribution, whereas the nature of the dose rate measurement leads to an arithmetic average. This means that using the CAM to determine the equivalent dose will undoubtedly underestimate the true burial doses and age underestimation is likely to result. Guérin et al. (2017) developed the Average Dose Model (ADM) to address these shortcomings of the CAM. In their study they compare the results for CAM and ADM applied to dose distributions obtained from 19 samples with independent age control up to $\sim$ 45 ka. They found that using the CAM resulted in an average age underestimate of 8 ± 2 %, whereas the ADM on average recovered the independent age control (ratio of 0.99 ± 0.02). The application of the ADM requires knowledge of the intrinsic over-dispersion, i.e., the over-dispersion arising from intrinsic sources of uncertainty σ_m (Thomsen et al., 2005; Galbraith et al., 2005; Thomsen et al., 2012), which most readily can be determined from dose recovery experiments. Ideally, σ_m should be determined by giving a uniform dose (i.e., a gamma dose, Thomsen et al., 2005, 2007, 2012; Guérin et al., 2017), but here we have chosen to use a beta dose instead although we acknowledge that this may result in an underestimation of σ_m. In our beta dose recovery experiment we determined σ_m to be $36 \pm 8\%$.

The effect of applying the various rejection criteria schemes on the ADM dose compared to simply including all grains with $s_{T_n} < 20\%$ is shown in Figure 5.11B. The average ratio of SG Std. ADM dose-to-MG quartz arithmetic dose is 0.82 ± 0.03 (n = 12). Thus, using the ADM the dose estimate increases by about 4% compared to using CAM. Application of the D_c criterion reduces the difference between single-grain ADM and multi-grain quartz doses, but does not completely remove it (average ratio is 0.88 ± 0.03).

All SG Q measurements have also been analysed using BayLum (see Table 5.4). The average dose ratio of SG BayLum (including saturated grains in the analysis) and SG CAM is 1.32 ± 0.05 (n = 12). The average relative uncertainty is $3.8 \pm 0.4\%$ and $3.9 \pm 0.5\%$ for CAM and BayLum, respectively. Thus, in contrast to the MG results, we observe a significant difference between the standard frequentist approach (using CAM or ADM) and using the Bayesian approach for SG Q dose determination. Importantly, we find that the average ratio of SG BayLum to MG is 1.05 ± 0.04

(n = 12), i.e., the results are statistically indistinguishable from one another. For these SG BayLum results, we included grains not giving bounded estimates in the frequentist framework. To investigate if it matters whether these grains are included in the analysis or not, we reran BayLum without including them. We find an average $\text{BayLum}_{no\,sat}$-to-BayLum ratio of 0.910 ± 0.012, i.e., the inclusion of the saturated grains increases the dose estimate by about 10%.

5.7 K17 Ages

In Figure 5.12, we show both multi-grain quartz (average and BayLum ages), single-grain quartz BayLum ages and pIRIR_{225} feldspar fading corrected ages as a function of depth. Also shown is the tephra age control (dashed lines) and previously published calibrated ^{14}C ages. In Table 5.5 we give the multi-grain BayLum ages for quartz and average feldspar (fading corrected) ages. Given the good agreement between MG Q arithmetic average doses and the BayLum MG and SG doses, we only give the BayLum MG Q ages in Table 5.5. Unfortunately, we were not able to recover K-rich feldspar from 13 of the 40 samples, and only very little from the remaining 27 samples. We did not attempt to run BayLum on the KF doses, because of the relative low number of dose estimates. So for KF, we present standard arithmetic average, fading corrected ages. Fading correction increases the IR_{50} ages by ∼33 % ($g_{2days} = 2.79 \pm 0.11$ %/decade) and the pIRIR_{225} ages by ∼5 % ($g_{2days} = 0.56 \pm 0.03$ %/decade). In the previous section (section 5.6), we found a significant discrepancy between CAM SG doses and MG doses, i.e., SG doses underestimated MG doses by about 20% on average. An apparent underestimation of this magnitude has been reported previously (e.g., Thomsen et al., 2016; Singh et al., 2017; Guérin et al., 2017).

However, as stated above, it is often argued that SG measurements are more accurate than MG measurements and thus others may have preferred the SG Q ages to the MG Q ages. To investigate if the SG Q ages are likely to be more accurate, we can compare them to the independent age control provided by the tephra layer (at a depth of 548 cm) with an age of 39.85 ± 0.14 ka. The two samples (for which we have SG measurements) that brackets the tephra layer are samples 207737 (taken 90 cm above the tephra) and sample 207742 (taken 10 cm below the tephra). The SG CAM ages for these two samples are 26.4 ± 1.5 ka ($n = 216$) and 26.6 ± 1.5 ka

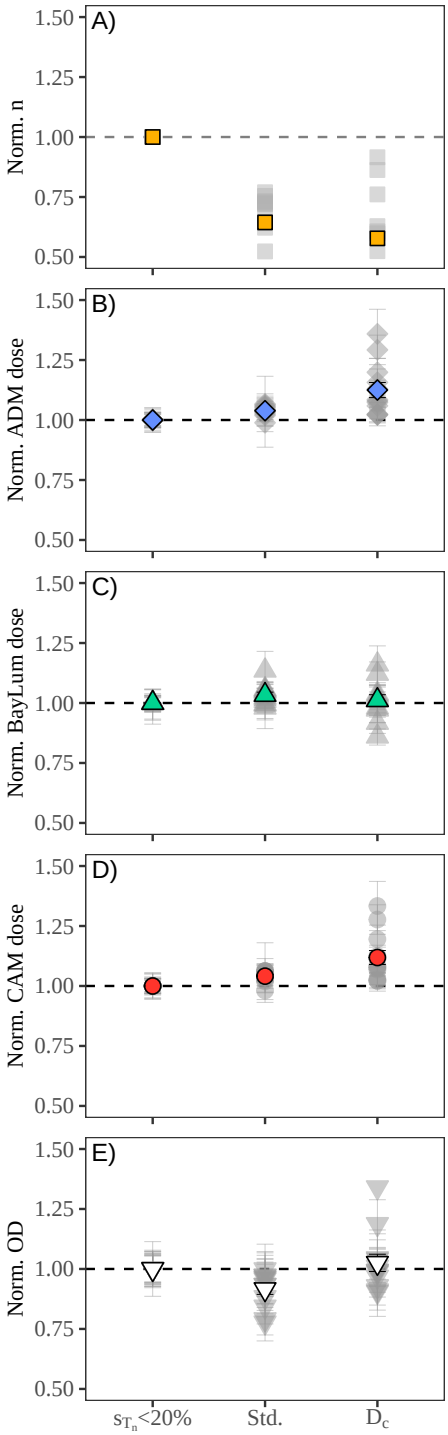

Figure 5.11: Effect of applying single-grain criteria relative to including all grains with $s_{T_n} < 20\%$ on the number of grains accepted into analysis of natural dose estimates (n), on Average Dose Model (ADM) dose estimates, on BayLum (including saturated grains) dose estimates, on Central Age Model (CAM) dose estimates and lastly on relative over-dispersion (OD). "Std." includes grains for which $s_{T_n} < 20\%$ recycling and IR depletion ratios are within 2σ of unity, and recuperation less than 5% of L_n/T_n. "D_c" only include grains with $s_{T_n} < 20\%$ and D_c values higher than the (running) CAM dose.

($n = 249$), respectively. Thus, the single-grain result for the deeper sample (207742) is underestimating the age control by $> 33\%$. The corresponding number for the MG result for this sample is $\sim 13\,\%$. Thus, it would appear that the standard MG Q measurements are considerably more accurate than the SG counterparts. It is however interesting to note that when using BayLum to estimate single-grain doses, we find a good agreement between our SG and MG results. Thus, in the following we will only discuss the MG results.

5.7.1 Multi-grain quartz BayLum age-depth profile

The two topmost samples (207752 and -53) sampled at less than 100 cm depth give MG Q BayLum ages whose 68% credible interval limits range between 7.4 and 9.1 (see Table 5.5). Then, just below a depth of 100 cm, the ages jump to 24.3 ka, 68% credible interval [22.7 ka, 25.6 ka]. The ages then increase systematically until a depth > 282 cm, where the ages suddenly become ~ 15 ka younger. The ages then increase systematically again with an age of the deepest sample of 53.8 ka, 68% credible interval [50.4 ka, 56.6 ka]. There is some scatter between the individual ages as is often seen in quartz OSL dating, but only 5 of the remaining 30 ages are inconsistent with the age immediately above or below it (i.e., samples 207736, 207742, 207743, 207745 and 207750).

In the following section, we consider the possibility that reworking and incomplete bleaching could be the cause of the age inversion of the samples collected between depths of 100 and 282 cm by discussing how the quartz and feldspar ages compare with each other.

5.7.2 Multi-grain quartz and feldspar comparison

For an OSL age to be accurate, it is important that there is no latent OSL signal from previous burials, as this would result in an age overestimation. One way to identify if the quartz signal was well-bleached at burial is to compare ages from quartz and feldspar (e.g., Murray et al., 2012). When exposed to the full daylight spectrum, the quartz OSL signal bleaches about one order of magnitude faster than the feldspar IR_{50} signal, which again bleaches significantly faster than the $pIRIR_{225}$ signal (Godfrey-Smith et al., 1988; Thomsen et al., 2008). If feldspar ages are "much" older, it could mean that both quartz and feldspar are incompletely bleached or that quartz was well-bleached and feldspar not. However, if the quartz and feldspar ages agree, we can confidently assume that both quartz and feldspar signals were bleached

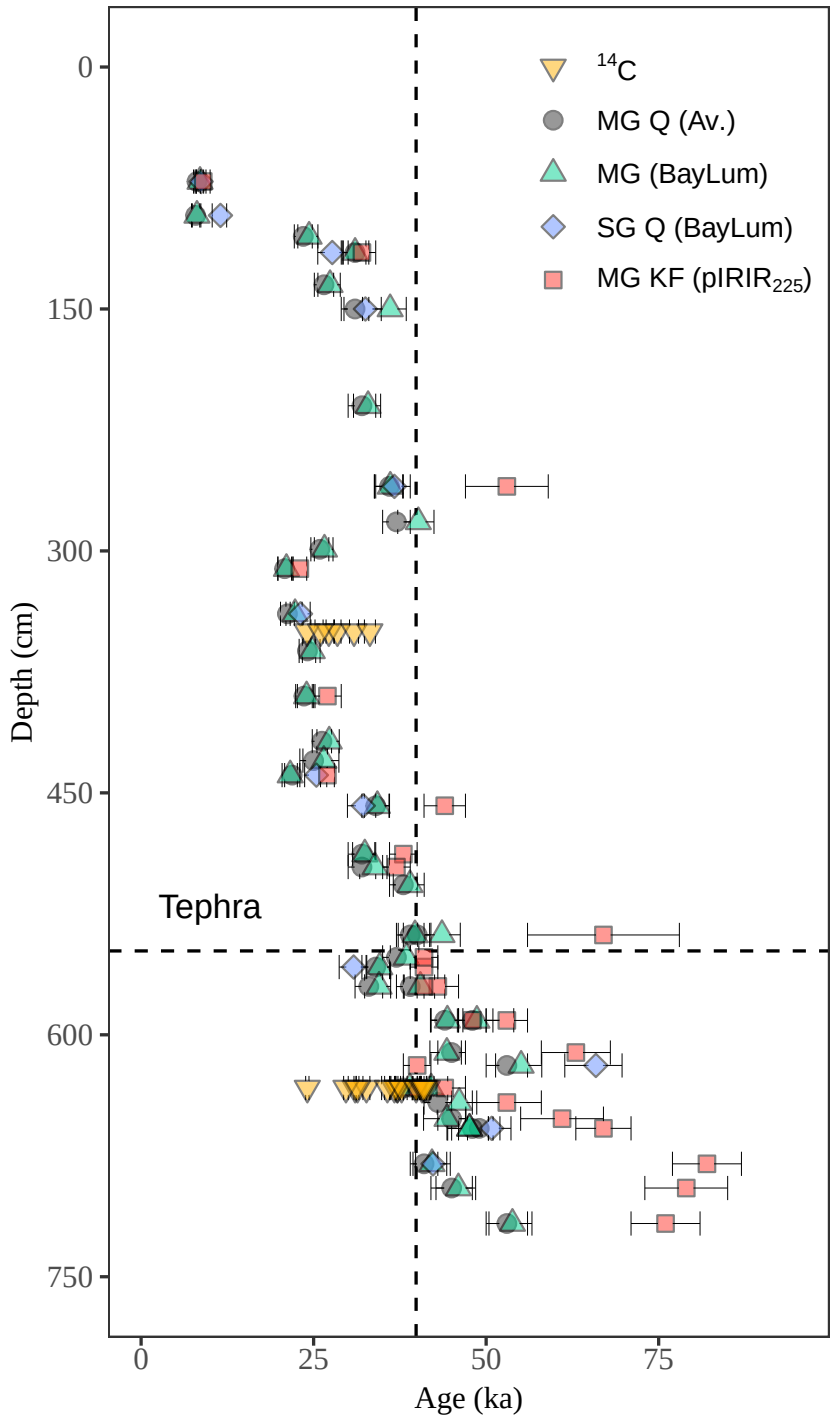

Figure 5.12: K17 depth-age profile showing multi-grain (MG) arithmetic average ages (Av., grey circles), MG Q BayLum ages (upward-facing green triangles), SG Q BayLum ages (blue diamonds) and fading corrected pIRIR$_{225}$ ages (red squares). Published radiocarbon ages (recalibrated here using Intcal20) are shown as downward facing light orange triangles. The depth (548 cm) and age (39.85 ± 0.14 ka) of the tephra horizon deposited by the Campanian Ignimbrite (CI) volcanic eruption are indicated by dashed lines. All ages are given at 68% confidence or credibility.

Table 5.5: K17 multi-grain (MG) quartz (Q) and feldspar (KF) ages. For MG Q the BayLum ages are given whereas for multi-grain feldspar, we show the fading-corrected IR_{50} (g_{2days} = 2.79 ± 0.11 %/decade) and $pIRIR_{225}$ (g_{2days} = 0.56 ± 0.03%/decade) arithmetic average ages. All ages are given at 68% confidence or credibility.

Lab code	Depth	MG Quartz age		MG Feldspar age			
	(cm)	n	BayLum (ka)	n	$pIRIR_{225}$ (ka)	n	IR_{50} (ka)
207752	71	24	8.5 [7.8, 9.1]	15	8.7 ± 1.4	14	7.2 ± 1.1
207753	92	20	8.1 [7.4, 8.7]				
207731	105	24	24.3 [22.7, 25.6]				
207732	115	24	31.0 [29.2, 32.6]	3	32 ± 2	3	24.7 ± 1.1
207733	135	24	27.4 [25.6, 28.9]				
207754	150	24	36.1 [32.1, 38.4]	3	126 ± 18	3	62 ± 3
207734	210	23	32.9 [30.8, 34.7]				
207735	260	23	36.1 [33.8, 37.9]	10	53 ± 6	11	53 ± 12
207758	282	24	40.2 [37.2, 42.4]				
207759	299	34	26.6 [25.1, 27.8]				
207760	311	21	21.1 [19.8, 22.0]	3	22.7 ± 1.2	3	24.0 ± 1.4
207761	339	22	22.3 [21.0, 23.3]				
207762	362	24	24.8 [23.4, 25.9]				
207764	390	22	24.0 [22.7, 25.2]	5	27 ± 2	5	26 ± 3
207765	418	24	27.3 [25.5, 28.7]				
207766	430	10	26.5 [23.5, 28.7]				
207736	439	36	21.6 [20.4, 22.6]	11	27 ± 1.3	12	25 ± 1.4
207737	458	22	34.2 [32.3, 35.9]	11	44 ± 3	11	37 ± 4
207738	488	23	32.4 [30.7, 33.9]	6	38 ± 2	6	31 ± 2
207739	496	24	33.9 [31.7, 35.6]	5	37 ± 2	5	34 ± 3
207740	507	24	39.0 [36.5, 41.0]				
207741	538	24	43.6 [40.1, 46.3]				
217701	538	23	39.7 [37.3, 41.8]	4	67 ± 11	4	45 ± 4
217702	552	24	38.4 [36.1, 40.3]	3	41 ± 2	3	36 ± 2
207742	558	34	34.6 [32.7, 36.2]	7	41 ± 2	7	36 ± 2
207743	570	36	34.5 [32.4, 36.2]	3	43 ± 3	3	36 ± 3
217703	570	24	40.5 [38.1, 42.5]	4	41 ± 3	4	37 ± 2
207744	591	24	44.4 [41.9, 46.6]	16	48 ± 6	16	37 ± 3
217704	591	37	48.7 [45.8, 51.0]	11	53 ± 3	11	44 ± 3
217705	611	24	44.3 [41.9, 46.4]	14	63 ± 5	14	57 ± 4
207745	619	24	55.1 [51.4, 58.1]	6	40 ± 2	6	35 ± 2
207747	633	70	38.9 [36.9, 40.6]	10	44 ± 3	10	40 ± 3
217706	633	32	42.0 [39.3, 44.4]	3	40 ± 2	3	37 ± 2
217707	642	33	46.1 [42.9, 48.6]	3	53 ± 5	3	51 ± 7
207748	652	23	44.3 [40.9, 47.1]	6	61 ± 6	6	48 ± 4
207749	658	24	47.6 [44.3, 50.3]	8	67 ± 4	8	52 ± 3
217708	658	30	47.6 [44.4, 50.4]				
207750	680	36	42.2 [39.8, 44.3]	11	82 ± 5	11	52 ± 3
217709	695	28	46.0 [42.7, 48.5]	16	79 ± 6	15	52 ± 3
217710	717	33	53.8 [50.4, 56.6]	9	76 ± 5	9	55 ± 4

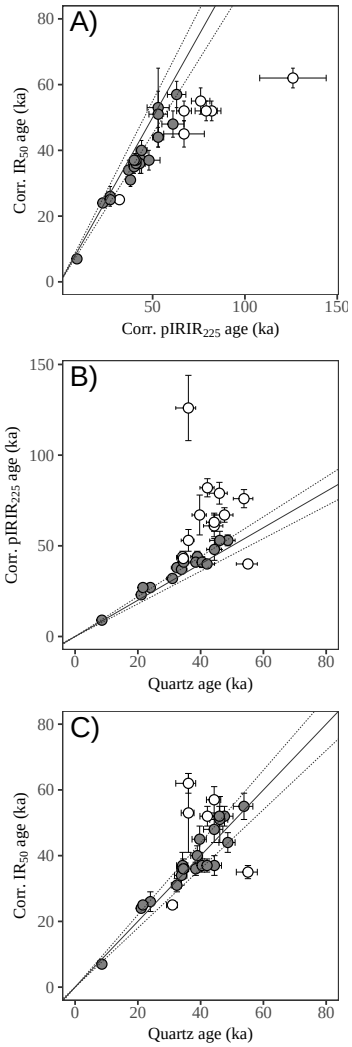

Figure 5.13: Comparison of multi-grain quartz BayLum ages and fading-corrected feldspar ages. A) $pIRIR_{225}$ against IR_{50}, B) $pIRIR_{225}$ against quartz, and C) IR_{50} against quartz. The 1:1 line is shown as a solid line and dashed lines indicate ± 10% of unity. Solid symbols represent sample ages consistent within 1σ of 1 ± 0.1. Open symbols represent those samples not satisfying this condition.

at burial. In Figure 5.13, we compare our quartz and feldspar ages. Also shown is the 1:1 line (solid line) and ± 10% of this line is indicated with dashed lines. In the following, we regard samples falling within 10 % of the 1:1 line (with uncertainties) as well-bleached (solid symbols) and those falling out as potentially poorly-bleached (open symbols).

In Figure 5.13A we plot the fading corrected feldspar IR_{50} ages against the fading corrected $pIRIR_{225}$ ages. The average ratio of IR_{50} to $pIRIR_{225}$ is 0.84 ± 0.12 (n = 27). For the young ages ($\lesssim$ 30 ka) there is a relatively good agreement between the two feldspar ages, but as the ages increase there is a tendency for the $pIRIR_{225}$

ages to be older than the IR_{50} ages. For ages $\gtrsim$ 60 ka about 50% of the $pIRIR_{225}$ ages significantly overestimate the IR_{50} ages. This could indicate, that in this age range, the $pIRIR_{225}$ signal from these samples was less well-bleached than the IR_{50} signal at burial, but given the systematic trend it could also indicate that our fading correction for the IR_{50} signal is too small. However, if we exclude the outliers as defined above the average ratio of IR_{50} to $pIRIR_{225}$ is only 0.89 ± 0.03 (n = 20).

In Figure 5.13B we plot the fading corrected $pIRIR_{225}$ ages against the quartz ages. The average ratio of $pIRIR_{225}$ to quartz is 1.32 ± 0.10 (n = 27). For the younger samples, there is a good agreement between the ages, but at ages $\gtrsim$ 35 ka ten (50%) of $pIRIR_{225}$ are significantly larger than the quartz ages. Excluding these samples the average ratio of $pIRIR_{225}$ to quartz ages is 1.10 ± 0.02 (n = 15).

In Figure 5.13C we plot the fading corrected feldspar IR_{50} ages against the quartz ages. The average ratio of IR_{50} to quartz is 1.06 ± 0.04 (n = 27). Of these, four samples show evidence of incomplete bleaching of the IR_{50} signal, while two IR_{50} ages significantly underestimate the corresponding quartz ages. Excluding these samples, the average ratio of IR_{50} to quartz is 1.02 ± 0.02 (n = 21). We conclude that the quartz signals were likely completely bleached at burial, because quartz ages are in agreement with IR_{50} ages - a signal that bleaches much slower than the quartz signal. We also observe that the $pIRIR_{225}$ of the older samples presumably suffer from a significant degree of incomplete bleaching at burial.

5.7.3 Descriptions of applied age models

To obtain a more accurate and precise estimate of the ages of the different culture layers and the CI/Y5 tephra layer, we model the MG Q age-depth relationship using three different Bayesian models: (i) *ArchaeoPhases* (Philippe and Vibet, 2020), (ii) *Bacon* (Blaauw and Christen, 2011a) and (iii) *OxCal* v4.4 (Bronk Ramsey, 2009). Our goal is to test each model and determine which, if any, is more appropriate for our data.

The "rBacon" R-package contains within it the "Bacon" model, a name "*...partly inspired by how specific prior information will produce smooth "floppy" or "crispy" Bayesian accumulation models*" (Blaauw and Christen, 2011a). Bacon works by dividing a sediment core into multiple sections of equal size, or thickness (Bacon()-function argument: "thick"), and then assuming that sediment accumulation hap-

pens linearly with time within each section. The goal is then to construct a model of sediment accumulation from which an age-depth relationship can be extracted. As previously mentioned, Bacon is a Bayesian model, and requires the input of several priors. One such prior exists for accumulation, which takes the form of a gamma distribution, whose mean and shape parameters can be set by the user ("acc.mean" and "acc.shape" arguments respectively). Adjacent section accumulation rates influences each other within Bacon, the degree of which is informed by a binomial distribution prior, whose arguments are set with "mem.mean" and "mem.strength" within the modelling function.

***ArchaeoPhases* modelling details**

The *ArchaeoPhases* "AgeDepth" function makes use of a matrix of ages, which is generated in a MCMC sampling procedure along with the depth of each sample. In this matrix, each column represents the age values sampled for one OSL sample. Here, we took the MCMC-matrix output of the multi-grain BayLum analysis, and created a MCMC-matrix subset, containing only the age-parameter columns. The fact that (currently) no other prior information can be input into the function to further constrain the model, means that this model is easy to run and reproduce.

Since *ArchaeoPhases* requires as input a MCMC-matrix, such as the one produced by BayLum, we also make use of the multi-grain BayLum ages for both the *Bacon* and *OxCal* models to allow a direct comparison of the three models. Uncertainties generated by BayLum are not necessarily symmetric around the most probable age value, but both *Bacon* and *OxCal* only allow input of a single value of uncertainty - not a range. Therefore we use the average of the distance from the interval limits to the central estimate.

***Bacon* modelling details**

For the "*Bacon*" modelling (where the sample depths of the individual samples are used as priors), we took a simple approach and kept most default values in the "*Bacon*()" function of the "r*Bacon*" R-package (Blaauw and Christen, 2011b). These defaults are "thick" = 5, "acc.shape" = 1.5, "acc.mean" = 20, "mem.strength" = 10, "mem.mean" = 0.5 and added no boundaries or hiatuses. Depending on the subset of samples modelled, the function will propose a different value of "acc.mean" which we accepted in all cases.

***OxCal* modelling details**

For *OxCal* age-modelling, we use the "P_Sequence" deposition model, which is a flexible compromise between using a model in which ages must be in a particular order ("Sequence()") and a model in which the sediment deposition rate is assumed to be uniform ("U_Sequence", Ramsey, 2006). The code-view input can be seen in Listing 5.1.

```
P_Sequence(1)
 {
 Boundary();
Date("s217710", N(2021-53668,3116)){z=717;};
 Date("s207750", N(2020-42063,2259)){z=680;};
 Date("s217707", N(2021-45962,2870)){z=642;};
 Date("s217705", N(2021-44229,2278)){z=611;};
 Date("s207743", N(2020-34403,1886)){z=570;};
 Date("s207741", N(2020-43392,3083)){z=538;};
 Date("s207738", N(2020-32351,1605)){z=488;};
 Date("s207737", N(2020-34158,1789)){z=458;};
 Date("s207765", N(2020-27165,1614)){z=418;};
 Date("s207762", N(2020-24739,1278)){z=362;};
 Date("s207761", N(2020-22265,1170)){z=339;};
 Date("s207759", N(2020-26514,1357)){z=299;};
Boundary();
};
```

Listing 5.1: *OxCal* Online input code to model an example set of K17 ages. "Date()" introduces an age set at a depth specified by "z"

5.7.4 Age-depth modelling results

From Figure 5.14A it is apparent that there is significant age inversion in the sediment column. From a depth of about 100 cm to about 285 cm, multi-grain quartz ages appear older than the layers immediately below. For age modelling purposes, we subdivided the full section into two subsections: One section from 100 cm and down to 282 cm (we designate this the "middle" section) and another section from 299 cm and downwards (we designate this the "lower" section). We omit the top two chernozem samples, as they may not represent single burial populations (see section 5.6.3). An exception is in our modelling of the full sequence (Figure 5.14A).

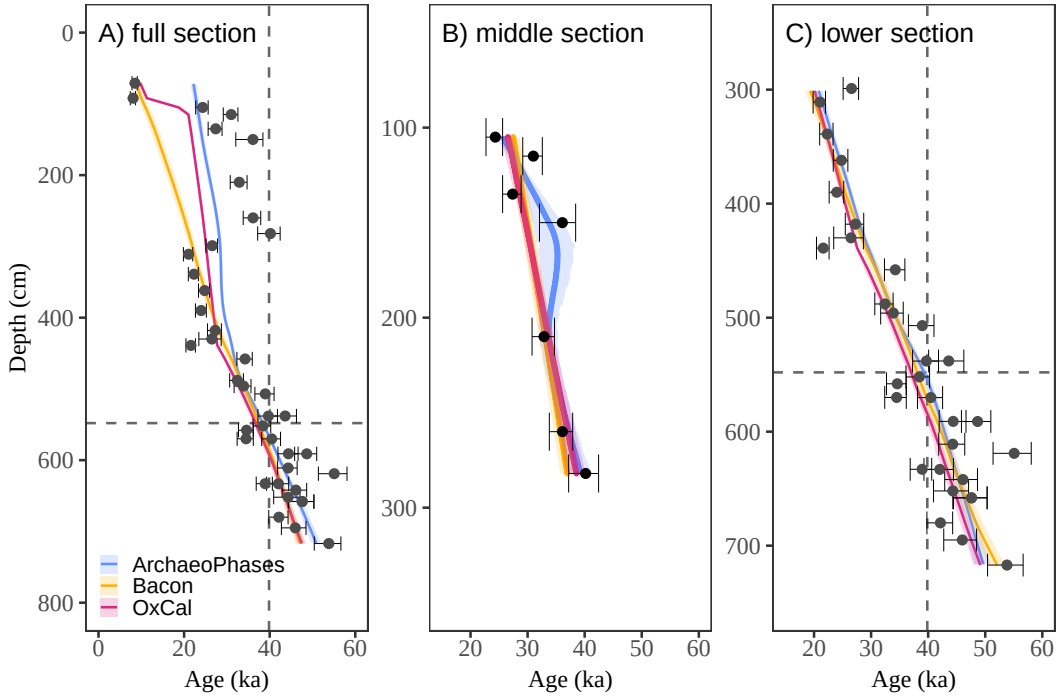

Figure 5.14: Age-depth models for the A) complete sequence, B) middle sequence, and C) lower sequence. We show the model results for the *ArchaeoPhases* age-depth model (blue), *Bacon* age-depth model (yellow) and *OxCal* 4.4 age-depth model (red). The different model estimates are represented as solid lines, while the associated bands show the 68 % probability range. Dashed lines indicate the depth and age of the tephra age control.

In what follows, it is important to reiterate that *Bacon* and *OxCal* allow for more complex modelling than attempted here. But in the simple forms employed here, it appears that *Bacon* creates a deposition rate (cm/ka) weighted heavily by the samples in the lower section, i.e., the deposition rate is more or less constant through out the full profile. This results in a continuous progression that is only slightly bent towards the outlying middle sequence. *OxCal*'s "P_Sequence()" model is not constrained by the assumption of a uniform deposition rate. In the full sequence, *ArchaeoPhases* does not capture the ages of the top chernozem samples very well. However, when running only the top and middle sequence together (not shown), *ArchaeoPhases* does in fact bend sharply inwards towards the top samples (as *OxCal* does). Overall, the three different age models yield similar results for the lower section, while there are significant differences in the middle section.

Given that modelling of the full sequence will be affected by a middle section which appear out of stratigraphic order (see section 5.8), we proceed with the "lower" modelling for age estimation of the archaeological layers and the tephra deposition (Table 5.6).

Table 5.6: Bayesian Age-model results based on multi-grain quartz BayLum ages from depths of 300-720 cm (the "lower section") at depths which correspond to Cultural layer 1, the Tephra layer and Cultural layer 2. The results are given at the 95 % credible interval.

Depth (cm)	Description	*ArchaeoPhases* (ka)	*Bacon* (ka)	*OxCal* (ka)
350	CL1a	23.8 [22.6, 24.9]	22.9 [21.4, 24.4]	22.7 [21.8, 23.6]
548	CI/Y5 tephra	39.3 [38.0, 40.4]	38.1 [36.4, 39.9]	37.0 [35.8, 38.1]
633	CL2a	44.8 [43.4, 46.1]	44.6 [42.7, 46.6]	43.2 [41.9, 44.5]
702	CL2b	48.9 [46.4, 51.4]	50.5 [47.9, 53.4]	48.0 [46.1, 50.0]

The CI/Y5 tephra layer provides a good independent age control at 39.85 ± 0.14 ka, that we can use to assess the different models. From Table 5.6, we see that our *ArchaeoPhases* age-depth model recovers the independent age control of the tephra, and actually does so within the 68% credible interval [38.7 ka, 39.9 ka] (see 5.15 and Table A.5), whereas the *Bacon* model recovers the age control within the 95% credible interval. However, our *OxCal* model does not recover the tephra age control within the 95% credible interval. Throughout the lower section, our OxCal model gives ages about 3% younger than our Bacon model, and 4% younger than our ArchaeoPhases model. Considering the results of all three models, the minimum and maximum 95% credible interval limits are 21.4 ka and 24.9 ka for CL1a, 41.9 ka and 46.6 ka for CL2a and 46.1 ka and 53.4 ka for CL2b.

5.7.5 Effect of sampling resolution on model estimates

In this study, OSL samples were taken with a median resolution of 16 cm/sample (duplicated samples not included), from a depth of 71 cm to 717 cm, for a total of 40 OSL samples (duplicated samples included). But what would we observe, had we sampled with a lower resolution? This question is important to answer given the observed scatter between individual quartz ages both in this study (Figure 5.14) and in general. Figure 5.15 shows *ArchaeoPhases*, *Bacon* and *OxCal* age model results, had we used only samples that best satisfy a target median sampling resolution. Starting with the top chernozem sample at a depth of 71 cm, we added samples closest to the depths specified by an arithmetic sequence of a target distance. To be able to make inferences about the age for CL2b, we ensured that bottom sample 217710 (depth of 717 cm) was always present in the sequence. As a result we derived OSL sample sets for a median resolution of 22 cm/sample, 30 cm/sample and 40 cm/sample. Our aim is to test how the age model results depend on the sampling resolution.

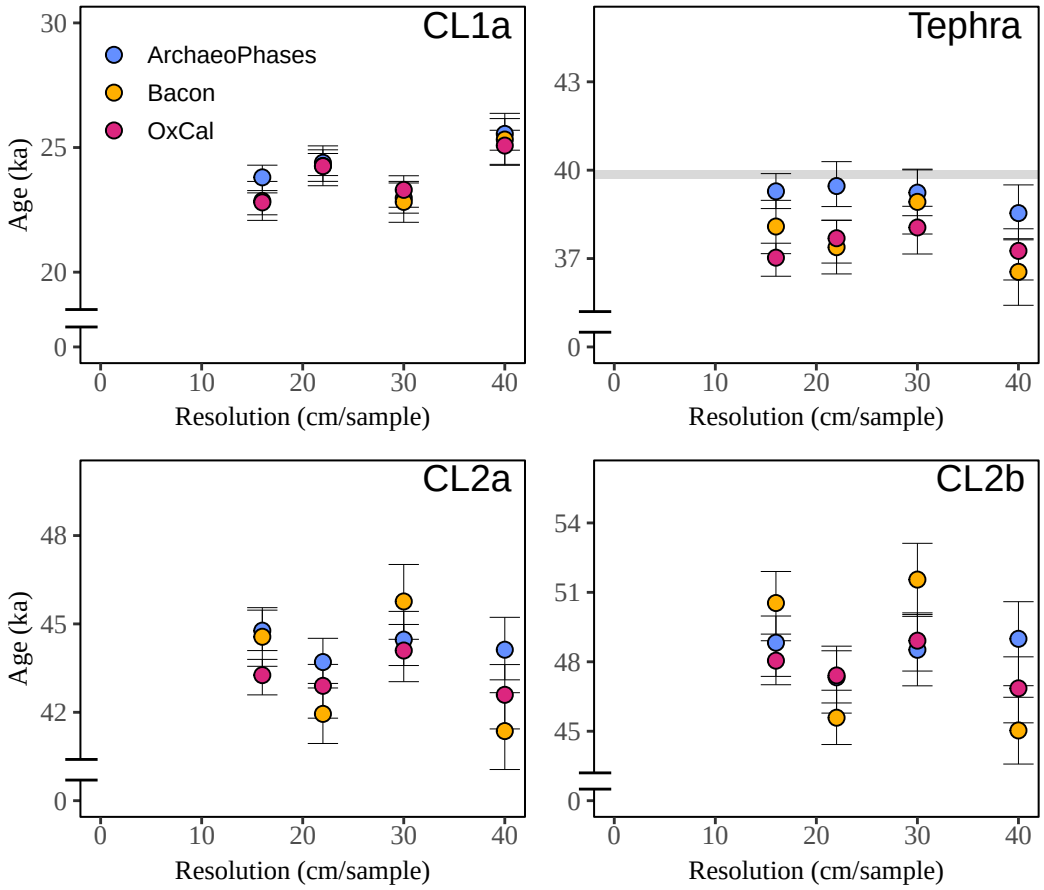

Figure 5.15: Effect of sampling resolution on age model results for the ages of the archaeological layers and the independent age control. The median vertical distance between samples for all K17 OSL samples is 16 cm per sample. We here show *ArchaeoPhases* (blue), *Bacon* (yellow) and *OxCal* (red) age-model results for CL1, the CI/Y5 tephra, CL2a and CL2b as a function of varying median distance between samples (resolution). The set of samples start with sample 207752 at a depth of 71 cm. Sequences are then filled with samples closest to the given resolution. For the actual modelling, any member of the outlying middle section was removed. Error bars indicate the limits of the 68% credible interval (see Figure 5.6).

In general, there appears to be little variation in the age estimate as a function of sampling resolution for a given age model (Figure 5.15).

In the following, we refer to two different comparisons of dispersions to describe the central age estimates shown in Figure 5.15. First, we investigate how dispersed the central estimates are for a given model for the different sampling resolutions (intra-model standard deviation, sd_{intra}). Then we compare the dispersion of age estimates between the different age models at each sampling resolution (inter-model standard deviation, sd_{inter}).

CL1a For CL1a (Figure 5.15A), sd_{intra} is 1.1 ka, 1.2 ka and 1.0 ka for *ArchaeoPhases*, *Bacon* and *OxCal*, respectively and sd_{inter} range between 0.1 and 0.6 ka. Thus,

for a given sampling resolution, the models return similar age estimates, but disperse more within themselves between sampling resolutions.

CI/Y5 tephra For the tephra layer (Figure 5.15B), sd_{intra} is lower: 0.4 ka (*ArchaeoPhases*), 1.0 ka (*Bacon*) and 0.6 ka (*OxCal*), but sd_{inter} is higher: 1.1 ka (res = 16 cm/sample), 1.1 ka (res = 22 cm/sample), 0.6 (res = 30 cm/sample) and 1.0 ka (res = 40 cm/sample).

CL2a For CL2a (Figure 5.15C), sd_{intra} is 0.6 ka (*ArchaeoPhases*), 2.1 ka (*Bacon*) and 0.6 ka (*OxCal*). The sd_{inter} is ∼0.8 ka for sampling resolutions of 16, 22 and 30 cm/sample, but with a sampling resolution of 40 cm/sample sd_{inter} increases to 1.4 ka.

CL2b CL2b (Figure 5.15D) mirrors the trends of CL2a.

The key observation from these comparisons is that at these four depths, our *ArchaeoPhases* model is least affected by sampling resolution (average sd_{intra} = 0.7 ka), whereas our *Bacon* model is most affected (average sd_{inter} = 1.9 ka).

5.8 Discussion

The multi-grain quartz age profile is characterized by a section of multi-grain OSL ages from 100 cm to 282 cm that appear older than expected when compared to the remaining profile, and the question of why begs examination. Stratigraphically, from roughly 200 cm and down to about 250 cm, a whitish layer consisting of clasts of limestone and sand appears. At this point in time in the burial history of K17, during heavy rains, the part of the gully in which K17 is situated was filled with sediment from above - eroded from cretaceous sediments from the upper part of gully. The ensuing slope movement of moisture-laden material, coupled with bioturbodation, caused these limestones and sands to mix with layers above and below. If these sediment materials were not sufficiently bleached, it would explain why we overestimate the age in this part of the column. We can at least partly test this hypothesis by comparing our quartz and feldspar ($pIRIR_{225}$) ages. Unfortunately, we only have three feldspar ages for this middle section (samples 207732, 207754 and 207735, see Table 5.5), which give feldspar to quartz ratios of 1.03 ± 0.09, 3.5 ± 0.6 and 1.5 ± 0.2, respectively. Thus, of the three samples, two might show evidence of some incomplete bleaching of at least the feldspar signal. But

for the one remaining sample (207732) which has an acceptable feldspar to quartz ratio, the quartz age is similar to the other two and all three are in stratigraphic order. Thus, we retain that the quartz signal in all likelihood was well-bleached at deposition. Nevertheless, the ages of the samples taken from below 285 cm increase continuously, and it is in these layers that both cultural layers reside, as well as the tephra.

When applying *ArchaeoPhases*, *Bacon* and *OxCal* age-depth models on this lower section, we generally derive similar age estimates. *ArchaeoPhases* recovers the independent age control of 39.85 ± 0.14 ka for the CI/Y5 tephra with a 68 % credible interval between 38.7 ka and 39.9 ka. Our *Bacon* model recovers the CI/Y5 tephra age control within the 95% credible interval, while our *OxCal* model underestimates the age control at the 95 % credible interval from 35.8 ka to 38.1 ka. Given that we compare the results of three separate age models, which model estimates should we trust more? Here, we emphasize our *ArchaeoPhases* age-depth model, since it more accurately recovers the CI/Y5 tephra (^{40}Ar/^{39}Ar) independent age control. In the previous subsection, we also showed that our ArchaeoPhases age-depth model is also the most robust model, compared to OxCal and Bacon. Hence, for the remainder of this chapter, we refer to OSL age results from our *ArchaeoPhases* age-depth model (The ArchaeoPhases K17 age chronology for the lower section is given in full in Tables A.1-A.5)

How does our chronology, based on multi-grain quartz OSL ages, compare with previously published ^{14}C ages for CL1 and CL2? It is clear that our OSL ages are significantly different from the radiocarbon ages previously published.
For CL1, Sinitsyn et al. (1997) provides a series of ^{14}C ages where the youngest age estimate is 27.2 to 24.6 kcal BP at 95% confidence (GIN-8076, Table 5.1), which is just consistent with our predicted age of 23.8 ka, 95% credible interval [22.6 ka, 24.9 ka]. However, the most recent radiocarbon dates from CL1 was published in 2019, and those results do not narrow the gap between the ^{14}C ages and the OSL ages presented here, but rather, they widen it. Dinnis et al. (2019a) dated a mammoth bone from CL1 to between 34.2 and 32.0 kcal BP (95% confidence), which is ∼10 ka (or almost 50 %) older than the OSL age.
In the case of CL2, then the most recent radiocarbon ages (Khlopachev, 2016; Dinnis et al., 2019a) range from 42.1 (OxA-X-2677-57) to 29.2 (OxA-30825) kcal BP. The

oldest ^{14}C age is dated to between 42.1 and 39.8 kcal BP (OxA-X-2677-57 Dinnis et al., 2019a) and is thus about 4 ka younger than our OSL estimate for CL2a of 44.8 ka, 95% credible interval [43.4 ka, 46.1 ka] and about 7-8 ka younger than our estimate for CL2b of 48.8 ka, 95% credible interval [46.4 ka, and 51.4 ka]. Our findings reveal the earliest Upper Palaeolithic evidence of AMH in Eastern Europe.

5.9 Conclusions

In this study we have derived OSL ages from 40 samples collected from a 7 m deep sediment column at the important Palaeolithic Kostenki 17 site in Russia.

We found that the quartz OSL signals are fast-component dominated and using a 260 ℃/ 220 ℃ preheat/cutheat SAR protocol recovers doses given in the laboratory prior to any thermal treatment. Using this protocol we measured the multi-grain quartz natural doses from the 40 OSL samples. We also measured the doses in 27 K-rich feldspar extracts using the feldspar pIRIR$_{225}$ signal ($g_{2days} = 0.56 \pm 0.03$%/decade).

Although quartz, IR$_{50}$ and pIRIR$_{225}$ age comparisons reveal that a significant fraction of the feldspar samples may suffer from incomplete bleaching of the pIRIR$_{225}$ signal, the quartz samples were most likely well-bleached at burial. We also measured 12 of the quartz samples using single-grain techniques and show that using the popular Central Age Model (CAM) to determine the burial dose results in an underestimation of the multi-grain doses by $\sim$ 20%. However, using a Bayesian approach to dose estimation (BayLum) in which individual grains, appearing to be in saturation using frequentist analysis, can be included, the single-grain dose estimates are indistinguishable from the multi-grain estimates.

The resulting age-depth profile (based on multi-grain quartz OSL ages) display some scatter between individual ages but most notably contain a section from a depth of 100 cm to 282 cm that currently inexplicably appears out of stratigraphic order. However, from a depth about 300 cm the remaining ages (n = 30) increase smoothly with depth. To obtain a more accurate and precise estimate of the ages of the different cultural layers and the CI/Y5 tephra layer, we used the output of multi-grain BayLum analysis and compared three age-depth Bayesian modelling procedures: (i) "AgeDepth()" of the R-package *ArchaeoPhases* (Philippe and Vibet, 2020), (ii) "Bacon()" of the R-package *rbacon* and (iii) "P_Sequence()" of *OxCal*. Our *ArchaeoPhases* age-depth model recovers the ^{40}Ar/^{39}Ar independent age con-

trol of the CI/Y5 tephra within a 68% credible interval and we thus consider it the best choice for our data. *ArchaeoPhases* predicts an age of 23.8 ka, 95% credible interval [22.6 ka, 24.9ka] for CL1a, an age of 44.8 ka, 95% credible interval [43.4 ka, 46.1 ka] for CL2a, and an age of 48.8 ka, 95% credible interval [46.4 ka, and 51.4 ka] for CL2b. Our multi-grain OSL ages differ significantly from previously published radiocarbon ages for both cultural layers, however given our excellent agreement with the $^{40}Ar/^{39}Ar$ tie point, we advocate that the OSL chronology presented here is more accurate. Our age for CL2 of 48.8 ka, 95% credible interval [46.4 ka, 51.4 ka], is evidence that the archaeological finds in this culture layer represent the earliest evidence of Anatomically Modern Humans (AMH) in this region of the world.

6 Summary, conclusions and outlook

Luminescence dating is an incredibly useful tool for a wide array of archaeological, geological and palaeo-climatological projects. Recent developments within the field of Luminescence has led to the creation of *BayLum* - a central dose and age model which takes the usual steps of analysis (from constructing dose response curves to calculating an age), and umbrellas them within a single Bayesian hierarchical model. In the process of using BayLum, I developed questions regarding both general and practical aspects of using BayLum. Through experiments with both real and simulated data sets, I found that the expected number of iterations to reach convergence of MCMC chains increased significantly with the number of samples included into the BayLum model. According to my research, this effect can be mitigated to some degree, by specifying suitable age-priors and by including a large number of measured units (i.e., aliquots and/or grains). Coupled with the features I implemented in BayLum, my work has significant implications for the application of BayLum. This is true generally, but especially when a large number of samples need to be included in the modelling and individual runs for each sample are not desirable.

Multi-grain (MG) and single-grain (SG) OSL dating of quartz are techniques with a growing catalogue of luminescence dating methodologies. While quartz SG dating is sometimes favored over quartz MG dating, it has been tested the least against independent age control, especially for samples older than 50 ka. In order to test common frequentist single-grain procedures, along with the new Bayesian one, I dated five Eemian samples (∼128 ka, 100-200 Gy dose range) and found that a multi-grain procedure, using a simple arithmetic average or BayLum, was the most accurate methodology, followed by single-grain BayLum (but only when saturated grains were included into the model). Both methods recovered, on average, the independent age-control. Conventional single-grain dating procedures significantly underestimated the age control in the 100-200 Gy dose range by almost 40-50%. These findings seriously question the accuracy of commonly applied single-grain

dating procedures in the 100-200 Gy dose range, but also show the promise of the BayLum approach.

One of my main objectives was to provide an OSL chronology for the Upper Palaeolithic site of Kostenki 17, Russia, where a cultural layer presumed to be of anatomically modern human (AMH) origin has been identified below a tephra layer produced by a volcanic eruption taking place ∼40 ka ago. Obtaining an accurate age for this cultural layer is an important piece of the puzzle in establishing when AMH first arrived in Eastern Europe. Radiocarbon dating has been applied on numerous occasions but is struggling to produce coherent results, not only for the cultural layer older than 40 ka but also for a younger cultural layer found above the tephra. The OSL chronology presented here has been obtained by using a suite of OSL techniques on both quartz and feldspar, but the final chronology has been obtained using BayLum on MG quartz measurements to produce individual OSL ages. Using Bayesian age-depth modelling software an age-depth model for Kostenki 17 was created. The cultural layer beneath the volcanic ash has an age of 48.9 ka, 95% credible interval [46.4 ka, 51.4 ka] and is thus the oldest known evidence of AMH in Upper Palaeolithic Eastern Europe.

This Ph.D. thesis highlights the promise of BayLum and Bayesian age-depth modelling software. But as to the future outlook of BayLum and the BayLum R-package, several avenues of research and development could be explored. One such avenue is to make BayLum more easy to use - both for the coding and non-coding population of luminescence researchers. One concrete example is to create/modify BayLum R-package functions such that they can work with imperfect binx-files. Another possible avenue is to explore BayLum features which I have not covered here, i.e., the ability to impose stratigraphical constraints on samples and the necessary development to make this type of modelling feasible for high-resolution sampling. In terms of the necessary development, one suggestion could be to explore the effect of setting good initial values for the MCMC procedure within BayLum. BayLum appears to bridge the gap between multi-grain and single-grain dating results in the 100-200 Gy dose range. However, the implications of the work presented here (although only based on five samples), is that common single-grain quartz procedures can fall significantly short in this age range. I believe these results warrant further investigations into the accuracy of single-grain dating with samples constrained at 50 ka or older,

where a significant portion of individual grains cannot record the absorbed dose due to saturation effects. Also, it questions if single-grain OSL dating really is required for older samples, when the standard and much more practical multi-grain approach appears to be more successful. In recent years, single-grain feldspar IRSL dating appears to be used more and more to derive OSL/IRSL chronologies. It would thus be relevant to investigate the accuracy of this approach using known-age samples. Given that the IRSL signal from K-rich feldspars grow to larger doses than quartz, it could be that feldspar does not suffer from the same shortcomings as quartz appear to do.

Bibliography

Aitken, M. J., Tite, M. S., and J, R. (1964). Thermoluminescent Dating of Ancient Ceramics. *Nature*, 202:1032–1033.

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.

Anikovich, M. (1992). Kul'tura verkhnego paleolita tsentral'noi i vostochnoi Evropy. *Arkheologicheskie Vesti*, 1:188–191.

Anikovich, M., Popov, V., and Platonova, N. (2008). Paleolit KostenkovskoBorshchevskogo Raiona v Kontekste Verkhnego Paleolita Evropy. Trudy KostenkovskoBorshchevskoi Arkheologocheskoi Ekspeditsii 1. Nestor-Istoriia, Saint Petersburg.

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., 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.

Arnold, L. J., Demuro, M., Parés, J. M., Pérez-González, A., Arsuaga, J. L., Bermúdez de Castro, J. M., and Carbonell, E. (2015). Evaluating the suitability of extended-range luminescence dating techniques over early and middle pleistocene timescales: Published datasets and case studies from atapuerca, spain. *Quaternary International*, 389:167–190. The Jaramillo Subchron and the Early-Middle Pleistocene transition in continental records from a multidisciplinary perspective.

Arnold, L. J. and Roberts, Richard, 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.

Auclair, M., Lamothe, M., and Huot, S. (2003). Measurement of anomalous fading for feldspar IRSL using SAR. *Radiation Measurements*, 37(4):487–492. Proceedings of the 10th international Conference on Luminescence and Electron-Spin Resonance Dating (LED 2002).

Auguie, B. (2019). *egg: Extensions for 'ggplot2': Custom Geom, Custom Themes, Plot Alignment, Labelled Panels, Symmetric Scales, and Fixed Panel Size*. R package version 0.4.5.

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.

Baumgarten, F. (2019). A morphometric exploration of subspecies taxonomy in a species of bird. Master's thesis, University of Southern Denmark.

Baumgarten, F., Philippe, A., Guillaume, G., and Kreutzer, S. (2023). Reducing computation time in the R-package'BayLum'. *Ancient TL*, 41(1):1–4.

Baumgarten, F., Thomsen, K. J., Guillaume, G., Buylaert, J.-P., and Murray, A. S. (Submitteda). Testing the accuracy of single-grain OSL dating on Eemian quartz samples. *Quaternary Geochronology*.

Baumgarten, F., Thomsen, K. J., Murray, A. S., Kurbanov, R. N., and Guérin, G. (to be submittedb). Establishing an OSL chronology for Kostenki 17 - an Upper Paleolithic site by the Don River, Russia. *to be decided*.

Bayliss, A. (2015). Quality in bayesian chronological models in archaeology. *World Archaeology*, 47:677–700.

Berger, G. W. (1995). Progress in luminescence dating methods for Quaternary sediments. *Dating methods for Quaternary deposits*, 2:81–104.

Blaauw, M. and Christen, J. A. (2011a). Flexible paleoclimate age-depth models using an autoregressive gamma process. *Bayesian Analysis*.

Blaauw, M. and Christen, J. A. (2011b). Flexible paleoclimate age-depth models using an autoregressive gamma process. *Bayesian Analysis*, 6(3):457 – 474.

Brock, F., Higham, T., Ditchfield, P., and Ramsey, C. (2010). Current pretreatment methods for ams radiocarbon dating at the oxford radiocarbon accelerator unit (orau). *Radiocarbon*, 52:103–112.

Bronk Ramsey, C. (2009). Bayesian analysis of radiocarbon dates. *Radiocarbon*, 51(1):337–360.

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. Proceedings of the 15th Solid State Dosimetry (SSD15).

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.

Buylaert, J.-P., Murray, A., Gebhardt, A., Sohbati, R., Ohlendorf, C., Thiel, C., Wastegård, S., Zolitschka, B., Team, P. S., et al. (2013). Luminescence dating of the PASADO core 5022-1D from Laguna Potrok Aike (Argentina) using IRSL signals from feldspar. *Quaternary Science Reviews*, 71:70–80.

Buylaert, J.-P., Murray, A. S., Thomsen, K. J., and Jain, M. (2009). Testing the potential of an elevated temperature IRSL signal from K-feldspar. *Radiation measurements*, 44(5-6):560–565.

Bøtter-Jensen, L., Andersen, C., Duller, G., and Murray, A. (2003a). Developments in radiation, stimulation and observation facilities in luminescence measurements.

Radiation Measurements, 37(4):535–541. Proceedings of the 10th international Conference on Luminescence and Electron-Spin Resonance Dating (LED 2002).

Bøtter-Jensen, L., Andersen, C., Duller, G., and Murray, A. (2003b). Developments in radiation, stimulation and observation facilities in luminescence measurements. *Radiation Measurements*, 37(4):535–541. Proceedings of the 10th international Conference on Luminescence and Electron-Spin Resonance Dating (LED 2002).

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.

Carter, T., Contreras, D. A., Holcomb, J., Mihailović, D. D., Karkanas, P., Guérin, G., Taffin, N., Athanasoulis, D., and Lahaye, C. (2019). Earliest occupation of the Central Aegean (Naxos), Greece: Implications for hominin and Homo sapiens' behavior and dispersals. *Science Advances*, 5(10):eaax0997.

Cherdyntsev, V., Alekseyev, V., Kind, N., Forova, V., Zavel'skiy, F., Sulerzhitskiy, L., and Forsenkova, I. (1968). Geological institute radiocarbon dates ii. *Radiocarbon*, 10:426–436.

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*.

Cresswell, A., Carter, J., and Sanderson, D. (2018). Dose rate conversion parameters: Assessment of nuclear data. *Radiation Measurements*, 120:195–201. 15th International Conference on Luminescence and Electron Spin Resonance Dating, 11-15 September 2017.

Daniels, F., Boyd, C., and Saunders, D. (1953). Thermoluminescence as a research tool. *Science*, 117(3040):343–349.

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. 15th International Conference on Luminescence and Electron Spin Resonance Dating, 11-15 September 2017, Cape Town, South Africa.

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.

Denwood, M. J. (2016). runjags: An R Package Providing Interface Utilities, Model Templates, Parallel Computing Methods and Additional Distributions for MCMC Models in JAGS. *Journal of Statistical Software*, 71:1–25.

Devièse, T., Abrams, G., Hajdinjak, M., Pirson, S., De Groote, I., Di Modica, K., Toussaint, M., Fischer, V., Comeskey, D., Spindler, L., et al. (2021). Reevaluating the timing of Neanderthal disappearance in Northwest Europe. *Proceedings of the National Academy of Sciences*, 118(12):e2022466118.

Deviese, T., Comeskey, D., McCullagh, J., Ramsey, C. B., and Higham, T. (2018). New protocol for compound-specific radiocarbon analysis of archaeological bones. *Rapid Communications in Mass Spectrometry*, 32(5):373–379.

Dietze, M. and Kreutzer, S. (2022). *sandbox: Probabilistic Numerical Modelling of Sediment Properties*. R package version 0.2.1.

Dinnis, R., Bessudnov, A., Artyushenko, A., Lada, A., Sinitsyn, A., and Higham,

T. (2019a). Kostënki 17 (Spitsynskaya) and Kostënki 6 (Streletskaya): recent fieldwork and new ^{14}C dates. *Quartär*, 66:225 – 230.

Dinnis, R., Bessudnov, A., Reynolds, N., Devièse, T., Pate, A., Sablin, M., Sinitsyn, A., and Higham, T. (2019b). New data for the Early Upper Paleolithic of Kostenki (Russia). *Journal of Human Evolution*, 127:21–40.

Duller, G. (1995). Luminescence dating using single aliquots: methods and applications. *Radiation Measurements*, 24(3):217–226.

Duller, G. (2003). Distinguishing quartz and feldspar in single grain luminescence measurements. *Radiation Measurements*, 37(2):161–165.

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

Duller, G., Bøtter-Jensen, L., Kohsiek, P., and Murray, A. (1999). A High-Sensitivity Optically Stimulated Luminescence Scanning System for Measurement of Single Sand-Sized Grains. *Radiation Protection Dosimetry*, 84.

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

Feathers, J. K., Rhodes, E. J., Huot, S., and Mcavoy, J. M. (2006). Luminescence dating of sand deposits related to late Pleistocene human occupation at the Cactus Hill Site, Virginia, USA. *Quaternary Geochronology*, 1(3):167–187.

Friedrich, J., Kreutzer, S., and Schmidt, C. (2022). *RLumModel: Solving Ordinary Differential Equations to Understand Luminescence*. R package version 0.2.10.

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. 15th International Conference on Luminescence and Electron Spin Resonance Dating, 11-15 September 2017, Cape Town, South Africa.

Fullagar, R., Price, D., and Head, L. (1996). Early human occupation of northern

Australia: Archaeology and thermoluminescence dating of Jinmium Rock Shelter, Northern Territory. *Top 100 Citations*, 70.

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., Roberts, R., and Yoshida, H. (2005). Error variation in osl palaeodose estimates from single aliquots of quartz: a factorial experiment. *Radiation Measurements*, 39(3):289–307.

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. LED14 Proceedings.

Gelman, A. and Rubin, D. B. (1992). Inference from Iterative Simulation Using Multiple Sequences. *Statistical Science*, 7:457–472.

Giaccio, B., Hajdas, I., Isaia, R., and Deino, A. (2017). High-precision 14C and 40Ar/39Ar dating of the Campanian Ignimbrite (Y-5) reconciles the time-scales of climatic-cultural processes at 40 ka. *Scientific Reports*, 7.

Godfrey-Smith, D., Huntley, D., and Chen, W.-H. (1988). Optical dating studies of quartz and feldspar sediment extracts. *Quaternary Science Reviews*, 7(3):373–380.

Guérin, G., Aldeias, V., Baumgarten, F., Goldberg, P., Gómez-Olivencia, A., Lahaye, C., Madelaine, S., Maureille, B., Philippe, A., Sandgathe, D., et al. (2023). A Third Neanderthal Individual from La Ferrassie Dated to the End of the Middle Palaeolithic. *PaleoAnthropology*, 2023(1):98–118.

Guérin, G., Lahaye, C., Heydari, M., Autzen, M., Buylaert, J.-P., Guibert, P., Jain, M., Kreutzer, S., Murray, A. S., Thomsen, K. J., et al. (2020). Towards an improvement of OSL age uncertainties: modelling OSL ages with systematic er-

rors, stratigraphic constraints and radiocarbon ages using the R package BayLum. *Geochronology Discussions*, 2020:1–30.

Guérin, G., Lebrun, B., Marchand, G., and Philippe, A. (2022). Age-depth modelling and the effect of including–or not–shared errors across sets of OSL samples: The case study of Beg-er-Vil (Brittany, France). *Quaternary Geochronology*, 70:101311.

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. 14th International Conference on Luminescence and Electron Spin Resonance Dating, 7–11 July, 2014, Montréal, Canada.

Guérin, G., Mercier, N., Nathan, R., Adamiec, G., and Lefrais, Y. (2012). On the use of the infinite matrix assumption and associated concepts: A critical review. *Radiation Measurements*, 47(9):778–785.

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.

Hawkins, D. M. (1980). *Identification of outliers*, volume 11. Springer.

Heydari, M., Guérin, G., Zeidi, M., and Conard, N. J. (2021). Bayesian luminescence dating at Ghār-e Boof, Iran, provides a new chronology for Middle and Upper Paleolithic in the southern Zagros. *Journal of human evolution*, 151:102926.

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.

Heydari, M., Guérin, G., Kreutzer, S., Jamet, G., Kharazian, M. A., Hashemi, M.,

Nasab, H. V., and Berillon, G. (2020). Do Bayesian methods lead to more precise chronologies? 'BayLum' and a first OSL-based chronology for the Palaeolithic open-air site of Mirak (Iran). *Quaternary Geochronology*, 59:101082.

Huntley, D. J., Godfrey-Smith, D. I., and Thewalt, M. L. W. (1985). Optical dating of sediments. *Nature*, 313:105–107.

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., Meyer, M., Roberts, R., Aldeias, V., Dibble, H., and El Hajraoui, M. (2011). Single-grain OSL dating at La Grotte des Contrebandiers ('Smugglers' Cave'), Morocco: improved age constraints for the Middle Paleolithic levels. *Journal of Archaeological Science*, 38(12):3631–3643.

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.

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.

Kaspar, F., Kühl, N., Cubasch, U., and Litt, T. (2005). A model-data comparison of european temperatures in the eemian interglacial. *Geophysical Research Letters*, 32(11).

Khlopachev, G. (2016). Upper Paleolithic. Images, symbols, signs. Catalog of small art objects from the archaeological collection of the MAE of RAS. St. Petersburg. In Russian.

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.

Kopp, R. E., Simons, F. J., Mitrovica, J. X., Maloof, A. C., and Oppenheimer, M. (2009). Probabilistic assessment of sea level during the last interglacial stage. *Nature*, 462:863–867.

Kozłowski, J. (1986). The Gravettian in Central and Eastern Europe. *Advances in World Archaeology*, 5:131–200.

Kreutzer, S., Burow, C., Dietze, M., Fuchs, M. C., Schmidt, C., Fischer, M., Friedrich, J., Mercier, N., Philippe, A., Riedesel, S., Autzen, M., Mittelstrass, D., Gray, H. J., and Galharret, J.-M. (2022). *Luminescence: Comprehensive Luminescence Dating Data Analysis*. R package version 0.9.20.

Kreutzer, S., Schmidt, C., Fuchs, M. C., Dietze, M., Fischer, M., and Fuchs, M. (2012). Introducing an R package for luminescence dating analysis. *Ancient TL*, 30(1):1–8.

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. LED14 Proceedings.

Kristensen, P. H., Gibbard, P. L., Knudsen, K. L., and Ehlers, J. (2000). Last interglacial stratigraphy at ristinge klint, south denmark. *Boreas*, 29:103–116.

Lahaye, C., Guérin, G., Gluchy, M., Hatté, C., Fontugne, M., Clemente-Conte, I., Santos, J. C., Villagran, X. S., Da Costa, A., Borges, C., et al. (2019). Another site, same old song: the Pleistocene-Holocene archaeological sequence of Toca da

Janela da Barra do Antonião-North, Piauí, Brazil. *Quaternary Geochronology*, 49:223–229.

Li, S.-H. (1994). Optical dating: insufficiently bleached sediments. *Radiation Measurements*, 23:563–567.

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.

Mahan, S. A., Rittenour, T. M., Nelson, M. S., Ataee, N., Brown, N., DeWitt, R., Durcan, J., Evans, M., Feathers, J., Frouin, M., et al. (2023). Guide for interpreting and reporting luminescence dating results. *Bulletin*, 135(5-6):1480–1502.

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.

McCulloch, M. T. and Esat, T. (2000). The coral record of last interglacial sea levels and sea surface temperatures. *Chemical Geology*, 169(1):107–129.

Medialdea, A., Thomsen, K., Murray, A., and Benito, G. (2014). Reliability of equivalent-dose determination and age-models in the OSL dating of historical and modern palaeoflood sediments. *Quaternary Geochronology*, 22:11–24.

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. 13th International Conference on Luminescence and Electron Spin Resonance Dating - LED 2011 Dedicated to J. Prescott and G. Berger.

Menke, B. (1985). Palynologische untersuchungen zur transgression des eem-meeres im raum offenbuttel/nord-ostsee-kanal. *Geo- . logisches Jahrbuch*, A86.

Monigal, K., Usik, V., Koulakovskaya, L., and Gerasimenko, N. (2006). The beginning of the Upper Paleolithic in Transcarpathia, Ukraine. *Anthropologie*, 44:61–74.

Murray, A. (1996). Developments in optically transferred luminescence and photo-transferred thermoluminescence dating: application to a 2000-year sequence of flood deposits. *Geochimica et Cosmochimica Acta*, 60:565–576.

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. LED 2002.

Murray, A., Marten, R., Johnston, A., and Martin, P. (1987). Analysis for naturally occuring 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. LED 2005.

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. Proceedings of the 13th International Conference on Luminescence and Electron Spin Resonance Dating, 10–14 July, 2011, Toruń, Poland.

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

Murray, A. S., Helsted, L. M., Autzen, M., Jain, M., and Buylaert, J.-P. (2018). Measurement of natural radioactivity: Calibration and performance of a high-resolution gamma spectrometry facility. *Radiation Measurements*, 120:215–220.

Murray, A. S., Olley, J. M., and Caitcheon, G. G. (1995). Measurement of equivalent doses in quartz from contemporary water-lain sediments using optically stimulated luminescence. *Quaternary Science Reviews*, 14:365–371.

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.

Philippe, A., Guerin, G., and Kreutzer, S. (2019). BayLum - An R package for Bayesian analysis of OSL ages: An introduction. *Quaternary Geochronology*, 49:16–24.

Philippe, A. and Vibet, M.-A. (2020). Analysis of archaeological phases using the R package ArchaeoPhases. *Journal of Statistical Software, Code Snippets*, 93(1):1–25.

Plummer, M. (2003). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. *Proceedings of the 3rd International Workshop on Distributed Statistical Computing, Vienna*, pages 1–10.

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.

Pyle, D. M., Ricketts, G. D., Margari, V., van Andel, T. H., Sinitsyn, A. A., Praslov, N. D., and Lisitsyn, S. (2006). Wide dispersal and deposition of distal tephra during the Pleistocene 'Campanian Ignimbrite/Y5' eruption, Italy. *Quaternary Science Reviews*, 25(21):2713–2728.

R Core Team (2023). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Ramsey, C. (2006). New approaches to constructing age models: OxCal4. *PAGES News*, 14:14–15.

Ramsey, C. B. (2008). Deposition models for chronological records. *Quaternary Science Reviews*, 27(1):42–60. INTegration of Ice-core, Marine and Terrestrial records (INTIMATE): Refining the record of the Last Glacial-Interglacial Transition.

Reimer, P. J., Austin, W. E. N., Bard, E., Bayliss, A., Blackwell, P. G., Bronk Ramsey, C., Butzin, M., Cheng, H., Edwards, R. L., Friedrich, M., and et al. (2020). The intcal20 northern hemisphere radiocarbon age calibration curve (0–55 cal kbp). *Radiocarbon*, 62(4):725–757.

Rhodes, E. (1988). Methodological considerations in the optical dating of quartz. *Quaternary Science Reviews*, 7(3):395–400.

Rhodes, E. J. and Pownall, L. (1994). Zeroing of the OSL signal in quartz from young glaciofluvial sediments. *Radiation Measurements*, 23:581–585.

Roberts, R., Bird, M., Olley, J., Galbraith, R., Lawson, E., Laslett, G., Yoshida, H., Jones, R., Fullagar, R., Jacobsen, G., and Hua, Q. (1998). Optical and radiocarbon dating at Jinmium rock shelter in northern Australia. *NATURE*, 393(6683):358–362.

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.

Sinitsyn, A., Praslov, N., Svezhentsev, Y., and Sulerzhitskiy, L. (1997). Radiocarbon chronology of the Upper Paleolithic of Eastern Europe. In: Sinitsyn, A.A., Praslov, N.D. (Eds.), Radiocarbon Chronology of the Paleolithic of Eastern Europe and Northern Asia. Problems and Perspectives, St. Petersburg.

Stepanova, K., Malutina, A., Bessudnov, A., and Girya, E. (2022). Personal ornaments from kostenki 17, layer ii: Manufacturing, usage and cultural context within the initial upper palaeolithic of eastern europe (in russian). *STRATUM PLUS*, pages 193–220.

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., 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., 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 Bøtter-Jensen, L. (2005). Sources of variability in OSL dose measurements using single grains of quartz. *Radiation measurements*, 39(1):47–61.

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

Thomsen, K. J., Murray, A. S., Jain, M., and Bøtter-Jensen, L. (2008). Laboratory fading rates of various luminescence signals from feldspar-rich sediment extracts. *Radiation Measurements*, 43:1474–1486.

Trachsel, M. and Telford, R. J. (2017). All age–depth models are wrong, but are getting better. *The Holocene*, 27(6):860–869.

Tveranger, J., Astakhov, V., and Mangerud, J. (1995). The Margin of the Last Barents-Kara Ice Sheet at Markhida, Northern Russia. *Quaternary Research*, 44(3):328–340.

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.

Wang, X., Lu, Y., and Wintle, A. (2006). Recuperated osl dating of fine-grained quartz in chinese loess. *Quaternary Geochronology*, 1(2):89–100.

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.

Wintle, A. and Murray, A. (1998). Towards the development of a preheat procedure for OSL dating of quartz. *Radiation Measurements*, 29(1):81–94.

Wintle, A. and Murray, A. (1999). Luminescence sensitivity changes in quartz. *Radiation Measurements*, 30(1):107–118.

Wintle, A. G. (2008). Fifty Years of Luminescence Dating*. *Archaeometry*, 50(2):276–312.

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., and Galbraith, R. (2000). Extending the age range of optical dating using single 'supergrains' of quartz. *Radiation Measurements*, 32(5-6):439–446.

Zhao, Q., Thomsen, K. J., Murray, A. S., Wei, M., Pan, B., Song, B., Zhou, R., Chen, S., Zhao, X., and Chen, H. (2015). Testing the use of OSL from quartz grains for dating debris flows in Miyun, northeast Beijing, China. *Quaternary Geochronology*, 30:320–327. LED14 Proceedings.

A Appendix I

Table A.1: Konstenki 17 OSL age chronology - Part 1. Age-depth model: Archaeophases. Only the lower section was included into the modelling.

Depth (cm)	Estimate (ka)	Credible interval (ka)				Depth (cm)	Estimate (ka)	Credible interval (ka)			
		$Lower_{68\%}$	$Upper_{68\%}$	$Lower_{95\%}$	$Upper_{95\%}$			$Lower_{68\%}$	$Upper_{68\%}$	$Lower_{95\%}$	$Upper_{95\%}$
300	20.8	20.1	21.4	19.5	22.1	342	23.3	22.7	23.8	22.3	24.5
301	20.8	20.2	21.5	19.5	22.1	343	23.4	22.8	23.9	22.4	24.5
302	20.9	20.2	21.5	19.6	22.2	344	23.4	22.9	23.9	22.4	24.5
303	20.9	20.3	21.6	19.7	22.2	345	23.5	22.9	24.0	22.5	24.6
304	21.0	20.4	21.7	19.8	22.3	346	23.6	23.0	24.0	22.6	24.7
305	21.1	20.4	21.7	19.8	22.3	347	23.6	23.1	24.1	22.7	24.7
306	21.1	20.5	21.8	19.9	22.4	348	23.7	23.1	24.2	22.7	24.8
307	21.2	20.6	21.8	20.0	22.5	349	23.7	23.2	24.2	22.8	24.9
308	21.3	20.6	21.9	20.0	22.5	350	23.8	23.3	24.3	22.9	24.9
309	21.3	20.7	22.0	20.1	22.6	351	23.9	23.3	24.4	22.8	24.9
310	21.4	20.8	22.0	20.2	22.6	352	23.9	23.4	24.4	22.9	24.9
311	21.4	20.8	22.1	20.3	22.7	353	24.0	23.5	24.5	23.0	25.0
312	21.5	20.9	22.1	20.3	22.8	354	24.1	23.6	24.6	23.1	25.1
313	21.6	21.0	22.2	20.4	22.8	355	24.1	23.6	24.6	23.1	25.1
314	21.6	21.0	22.2	20.4	22.9	356	24.2	23.7	24.7	23.2	25.2
315	21.7	21.1	22.3	20.5	22.9	357	24.3	23.7	24.7	23.3	25.3
316	21.7	21.1	22.4	20.6	23.0	358	24.3	23.8	24.8	23.3	25.3
317	21.8	21.2	22.4	20.6	23.0	359	24.4	23.9	24.9	23.4	25.4
318	21.9	21.3	22.5	20.7	23.1	360	24.5	24.0	25.0	23.5	25.5
319	21.9	21.3	22.5	20.8	23.1	361	24.5	24.1	25.0	23.5	25.5
320	22.0	21.3	22.5	20.8	23.2	362	24.6	24.1	25.1	23.6	25.6
321	22.0	21.4	22.6	20.9	23.2	363	24.7	24.2	25.2	23.7	25.6
322	22.1	21.5	22.7	21.0	23.3	364	24.7	24.3	25.2	23.8	25.7
323	22.2	21.5	22.7	20.9	23.3	365	24.8	24.3	25.3	23.8	25.8
324	22.2	21.7	22.8	21.0	23.3	366	24.9	24.4	25.3	23.9	25.8
325	22.3	21.6	22.8	21.1	23.4	367	24.9	24.5	25.4	24.0	25.9
326	22.3	21.7	22.8	21.2	23.5	368	25.0	24.6	25.5	24.1	26.0
327	22.4	21.7	22.9	21.3	23.6	369	25.1	24.6	25.6	24.1	26.0
328	22.5	21.8	23.0	21.4	23.6	370	25.1	24.7	25.6	24.2	26.1
329	22.5	21.9	23.0	21.4	23.7	371	25.2	24.7	25.7	24.3	26.1
330	22.6	21.9	23.1	21.5	23.7	372	25.2	24.8	25.7	24.3	26.2
331	22.6	22.0	23.1	21.5	23.8	373	25.3	24.9	25.8	24.4	26.3
332	22.7	22.1	23.2	21.6	23.8	374	25.4	24.9	25.9	24.5	26.3
333	22.8	22.2	23.3	21.6	23.8	375	25.4	25.0	25.9	24.5	26.4
334	22.8	22.3	23.4	21.7	23.9	376	25.5	25.1	26.0	24.6	26.4
335	22.9	22.3	23.4	21.9	24.1	377	25.6	25.1	26.0	24.6	26.5
336	22.9	22.4	23.5	21.9	24.1	378	25.6	25.2	26.1	24.7	26.5
337	23.0	22.4	23.5	22.0	24.2	379	25.7	25.2	26.1	24.7	26.6
338	23.1	22.5	23.6	22.1	24.2	380	25.7	25.2	26.1	24.8	26.6
339	23.1	22.6	23.6	22.1	24.3	381	25.8	25.3	26.2	24.9	26.7
340	23.2	22.6	23.7	22.2	24.3	382	25.9	25.4	26.3	24.9	26.7
341	23.2	22.7	23.8	22.3	24.4	383	25.9	25.4	26.3	25.0	26.8

Table A.2: Konstenki 17 OSL age chronology - Part 2. Age-depth model: Archaeophases. Only the lower section was included into the modelling.

Depth (cm)	Estimate (ka)	Credible interval (ka)				Depth (cm)	Estimate (ka)	Credible interval (ka)			
		$Lower_{68\%}$	$Upper_{68\%}$	$Lower_{95\%}$	$Upper_{95\%}$			$Lower_{68\%}$	$Upper_{68\%}$	$Lower_{95\%}$	$Upper_{95\%}$
384	26.0	25.5	26.4	25.0	26.9	426	28.5	28.1	29.0	27.7	29.5
385	26.0	25.5	26.4	25.1	26.9	427	28.6	28.2	29.1	27.7	29.5
386	26.1	25.6	26.5	25.2	27.0	428	28.7	28.2	29.1	27.8	29.6
387	26.2	25.7	26.6	25.2	27.0	429	28.8	28.3	29.2	27.9	29.7
388	26.2	25.8	26.6	25.3	27.1	430	28.8	28.4	29.3	27.9	29.8
389	26.3	25.8	26.7	25.3	27.1	431	28.9	28.5	29.3	28.0	29.8
390	26.3	25.9	26.7	25.4	27.2	432	29.0	28.5	29.4	28.1	29.9
391	26.4	25.9	26.8	25.4	27.2	433	29.1	28.6	29.5	28.2	30.0
392	26.4	26.0	26.9	25.5	27.3	434	29.1	28.7	29.6	28.3	30.1
393	26.5	26.0	26.9	25.6	27.4	435	29.2	28.8	29.7	28.3	30.2
394	26.6	26.1	27.0	25.7	27.5	436	29.3	28.8	29.7	28.4	30.2
395	26.6	26.1	27.0	25.7	27.5	437	29.4	28.9	29.8	28.5	30.3
396	26.7	26.2	27.1	25.8	27.6	438	29.4	29.0	29.9	28.6	30.4
397	26.7	26.3	27.1	25.8	27.6	439	29.5	29.1	30.0	28.6	30.5
398	26.8	26.3	27.2	25.9	27.7	440	29.6	29.2	30.1	28.7	30.5
399	26.8	26.4	27.3	25.9	27.7	441	29.7	29.3	30.1	28.8	30.6
400	26.9	26.4	27.3	26.0	27.8	442	29.8	29.3	30.2	28.9	30.7
401	27.0	26.5	27.4	26.1	27.9	443	29.8	29.4	30.3	29.0	30.8
402	27.0	26.6	27.4	26.1	27.9	444	29.9	29.5	30.4	29.0	30.9
403	27.1	26.6	27.5	26.2	28.0	445	30.0	29.6	30.5	29.1	30.9
404	27.1	26.7	27.6	26.2	28.0	446	30.1	29.7	30.6	29.2	31.0
405	27.2	26.8	27.6	26.3	28.1	447	30.2	29.7	30.6	29.3	31.1
406	27.3	26.8	27.7	26.3	28.1	448	30.2	29.8	30.7	29.3	31.1
407	27.3	26.9	27.7	26.4	28.2	449	30.3	29.9	30.8	29.4	31.2
408	27.4	27.0	27.8	26.5	28.3	450	30.4	30.0	30.9	29.5	31.3
409	27.4	27.0	27.9	26.5	28.3	451	30.5	30.0	30.9	29.6	31.4
410	27.5	27.1	27.9	26.6	28.4	452	30.5	30.1	31.0	29.6	31.5
411	27.6	27.1	28.0	26.6	28.4	453	30.6	30.2	31.1	29.7	31.6
412	27.6	27.2	28.1	26.7	28.5	454	30.7	30.3	31.2	29.8	31.6
413	27.7	27.2	28.1	26.8	28.6	455	30.8	30.3	31.3	29.9	31.7
414	27.7	27.3	28.2	26.8	28.6	456	30.8	30.4	31.4	30.0	31.8
415	27.8	27.4	28.3	26.9	28.7	457	30.9	30.5	31.4	30.0	31.9
416	27.9	27.4	28.3	27.0	28.8	458	31.0	30.6	31.5	30.1	32.0
417	27.9	27.5	28.4	27.1	28.9	459	31.1	30.6	31.6	30.2	32.0
418	28.0	27.5	28.4	27.1	28.9	460	31.1	30.7	31.7	30.3	32.1
419	28.1	27.6	28.5	27.2	29.0	461	31.2	30.8	31.7	30.3	32.2
420	28.1	27.7	28.6	27.3	29.1	462	31.3	30.9	31.8	30.4	32.3
421	28.2	27.8	28.6	27.3	29.1	463	31.4	30.9	31.9	30.5	32.4
422	28.3	27.8	28.7	27.4	29.2	464	31.5	31.0	32.0	30.6	32.5
423	28.3	27.9	28.8	27.5	29.3	465	31.5	31.1	32.1	30.6	32.5
424	28.4	28.0	28.8	27.5	29.3	466	31.6	31.2	32.1	30.7	32.6
425	28.5	28.0	28.9	27.6	29.4	467	31.7	31.2	32.2	30.8	32.7

Table A.3: Konstenki 17 OSL age chronology - Part 3. Age-depth model: Archaeophases. Only the lower section was included into the modelling.

Depth (cm)	Estimate (ka)	Credible interval (ka)				Depth (cm)	Estimate (ka)	Credible interval (ka)			
		$Lower_{68\%}$	$Upper_{68\%}$	$Lower_{95\%}$	$Upper_{95\%}$			$Lower_{68\%}$	$Upper_{68\%}$	$Lower_{95\%}$	$Upper_{95\%}$
468	31.8	31.3	32.3	30.9	32.8	510	35.4	34.8	35.9	34.4	36.5
469	31.8	31.3	32.3	30.9	32.9	511	35.5	34.9	36.0	34.5	36.6
470	31.9	31.5	32.5	31.0	33.0	512	35.6	35.1	36.1	34.6	36.7
471	32.0	31.5	32.5	31.1	33.1	513	35.7	35.2	36.2	34.7	36.8
472	32.1	31.6	32.6	31.1	33.1	514	35.8	35.3	36.3	34.8	36.9
473	32.1	31.7	32.7	31.2	33.2	515	35.9	35.3	36.4	34.9	37.1
474	32.2	31.8	32.8	31.3	33.3	516	36.0	35.4	36.5	35.0	37.1
475	32.3	31.8	32.9	31.4	33.4	517	36.1	35.5	36.6	35.1	37.3
476	32.4	31.9	32.9	31.4	33.5	518	36.2	35.6	36.7	35.2	37.4
477	32.4	32.0	33.0	31.5	33.5	519	36.3	35.7	36.8	35.3	37.5
478	32.5	32.1	33.1	31.6	33.6	520	36.4	35.8	36.9	35.4	37.6
479	32.6	32.1	33.2	31.6	33.7	521	36.5	35.9	37.0	35.5	37.7
480	32.7	32.2	33.2	31.7	33.8	522	36.6	36.1	37.2	35.6	37.8
481	32.8	32.3	33.3	31.8	33.9	523	36.7	36.2	37.3	35.7	37.9
482	32.8	32.4	33.4	31.9	33.9	524	36.8	36.3	37.4	35.8	38.0
483	32.9	32.4	33.5	31.9	34.0	525	37.0	36.4	37.5	35.8	38.1
484	33.0	32.5	33.6	32.0	34.1	526	37.1	36.5	37.6	36.0	38.2
485	33.1	32.6	33.7	32.1	34.2	527	37.2	36.6	37.7	36.0	38.2
486	33.2	32.7	33.7	32.2	34.3	528	37.3	36.7	37.8	36.2	38.4
487	33.3	32.8	33.8	32.2	34.3	529	37.4	36.8	37.9	36.3	38.5
488	33.3	32.9	33.9	32.3	34.4	530	37.5	36.9	38.0	36.4	38.6
489	33.4	32.9	34.0	32.4	34.5	531	37.6	37.0	38.1	36.5	38.7
490	33.5	33.0	34.1	32.4	34.6	532	37.7	37.1	38.2	36.6	38.9
491	33.6	33.1	34.1	32.5	34.6	533	37.8	37.2	38.3	36.6	38.9
492	33.7	33.2	34.2	32.6	34.7	534	37.9	37.3	38.4	36.7	39.0
493	33.8	33.2	34.3	32.7	34.8	535	38.0	37.4	38.5	36.8	39.1
494	33.9	33.3	34.4	32.8	34.9	536	38.1	37.5	38.6	36.9	39.2
495	33.9	33.4	34.5	32.9	35.0	537	38.2	37.6	38.7	37.0	39.3
496	34.0	33.5	34.6	33.0	35.1	538	38.3	37.7	38.8	37.1	39.4
497	34.1	33.6	34.7	33.1	35.2	539	38.4	37.8	38.9	37.2	39.5
498	34.2	33.7	34.8	33.1	35.3	540	38.5	37.9	39.0	37.3	39.7
499	34.3	33.8	34.9	33.2	35.4	541	38.6	38.0	39.1	37.4	39.7
500	34.4	33.8	34.9	33.3	35.5	542	38.7	38.1	39.2	37.5	39.8
501	34.5	33.9	35.0	33.4	35.6	543	38.8	38.2	39.4	37.6	39.9
502	34.6	34.0	35.1	33.6	35.7	544	38.9	38.3	39.5	37.6	40.0
503	34.7	34.1	35.2	33.6	35.8	545	39.0	38.3	39.5	37.8	40.2
504	34.8	34.2	35.3	33.8	35.9	546	39.1	38.5	39.7	37.8	40.2
505	34.9	34.3	35.4	33.9	36.0	547	39.2	38.6	39.8	38.0	40.4
506	35.0	34.4	35.5	34.0	36.1	548	39.3	38.7	39.9	38.0	40.4
507	35.1	34.5	35.6	34.1	36.2	549	39.4	38.8	39.9	38.1	40.5
508	35.2	34.6	35.7	34.2	36.3	550	39.5	38.8	40.0	38.3	40.7
509	35.3	34.7	35.8	34.3	36.4	551	39.5	38.9	40.1	38.4	40.8

Table A.4: Konstenki 17 OSL age chronology - Part 4. Age-depth model: Archaeophases. Only the lower section was included into the modelling.

Depth (cm)	Estimate (ka)	Credible interval (ka)				Depth (cm)	Estimate (ka)	Credible interval (ka)			
		$Lower_{68\%}$	$Upper_{68\%}$	$Lower_{95\%}$	$Upper_{95\%}$			$Lower_{68\%}$	$Upper_{68\%}$	$Lower_{95\%}$	$Upper_{95\%}$
552	39.6	39.0	40.2	38.5	40.9	594	42.5	41.9	43.1	41.3	43.7
553	39.7	39.1	40.3	38.6	41.0	595	42.6	41.9	43.1	41.4	43.8
554	39.8	39.2	40.4	38.6	41.0	596	42.6	42.0	43.2	41.5	43.9
555	39.9	39.2	40.4	38.8	41.2	597	42.7	42.1	43.3	41.5	43.9
556	40.0	39.3	40.5	38.8	41.2	598	42.7	42.2	43.4	41.6	44.0
557	40.1	39.4	40.6	38.9	41.3	599	42.8	42.3	43.5	41.6	44.0
558	40.1	39.5	40.7	39.0	41.4	600	42.9	42.3	43.5	41.7	44.1
559	40.2	39.6	40.8	39.0	41.4	601	42.9	42.4	43.6	41.7	44.1
560	40.3	39.7	40.9	39.2	41.5	602	43.0	42.4	43.6	41.8	44.2
561	40.4	39.8	40.9	39.2	41.5	603	43.0	42.5	43.7	41.9	44.2
562	40.4	39.9	41.0	39.3	41.6	604	43.1	42.5	43.7	41.9	44.3
563	40.5	39.9	41.0	39.3	41.7	605	43.2	42.6	43.8	41.9	44.3
564	40.6	39.9	41.1	39.4	41.7	606	43.2	42.6	43.8	42.1	44.5
565	40.6	40.0	41.1	39.5	41.8	607	43.3	42.7	43.9	42.1	44.5
566	40.7	40.1	41.2	39.5	41.9	608	43.3	42.7	43.9	42.2	44.6
567	40.8	40.2	41.3	39.7	42.0	609	43.4	42.8	44.0	42.2	44.6
568	40.9	40.3	41.5	39.7	42.1	610	43.4	42.8	44.0	42.3	44.7
569	40.9	40.4	41.5	39.8	42.2	611	43.5	42.9	44.1	42.4	44.7
570	41.0	40.4	41.6	39.9	42.2	612	43.6	42.9	44.1	42.4	44.8
571	41.1	40.4	41.6	40.0	42.3	613	43.6	42.9	44.2	42.5	44.8
572	41.1	40.5	41.7	40.0	42.4	614	43.7	43.0	44.2	42.5	44.9
573	41.2	40.6	41.7	40.1	42.4	615	43.7	43.1	44.3	42.5	44.9
574	41.2	40.6	41.8	40.1	42.5	616	43.8	43.1	44.4	42.6	45.0
575	41.3	40.7	41.9	40.2	42.6	617	43.8	43.2	44.5	42.6	45.0
576	41.4	40.8	41.9	40.2	42.6	618	43.9	43.3	44.6	42.7	45.1
577	41.4	40.8	42.0	40.3	42.6	619	44.0	43.3	44.6	42.7	45.2
578	41.5	40.9	42.1	40.3	42.7	620	44.0	43.4	44.7	42.7	45.2
579	41.6	41.0	42.2	40.4	42.8	621	44.1	43.4	44.7	42.8	45.3
580	41.6	41.1	42.2	40.5	42.8	622	44.1	43.5	44.8	42.9	45.4
581	41.7	41.1	42.3	40.5	42.9	623	44.2	43.6	44.9	43.0	45.4
582	41.8	41.2	42.4	40.6	42.9	624	44.2	43.6	44.9	43.0	45.5
583	41.8	41.2	42.4	40.7	43.0	625	44.3	43.7	45.0	43.0	45.5
584	41.9	41.3	42.5	40.7	43.1	626	44.4	43.7	45.0	43.1	45.6
585	41.9	41.4	42.5	40.8	43.2	627	44.4	43.8	45.1	43.2	45.7
586	42.0	41.4	42.6	40.9	43.2	628	44.5	43.8	45.2	43.2	45.8
587	42.1	41.5	42.6	40.9	43.3	629	44.5	43.9	45.2	43.3	45.8
588	42.1	41.5	42.7	41.0	43.4	630	44.6	43.9	45.3	43.3	45.9
589	42.2	41.6	42.8	41.0	43.4	631	44.7	44.0	45.3	43.4	45.9
590	42.3	41.6	42.8	41.1	43.5	632	44.7	44.0	45.4	43.4	46.0
591	42.3	41.7	42.9	41.2	43.5	633	44.8	44.1	45.5	43.4	46.1
592	42.4	41.8	42.9	41.2	43.6	634	44.8	44.1	45.5	43.5	46.2
593	42.4	41.8	43.0	41.3	43.7	635	44.9	44.2	45.6	43.6	46.3

Table A.5: Konstenki 17 OSL age chronology - Part 5. Age-depth model: Archaeophases. Only the lower section was included into the modelling.

Depth (cm)	Estimate (ka)	Credible interval (ka)				Depth (cm)	Estimate (ka)	Credible interval (ka)			
		$Lower_{68\%}$	$Upper_{68\%}$	$Lower_{95\%}$	$Upper_{95\%}$			$Lower_{68\%}$	$Upper_{68\%}$	$Lower_{95\%}$	$Upper_{95\%}$
636	44.9	44.3	45.7	43.6	46.3	678	47.5	46.4	48.5	45.5	49.5
637	45.0	44.3	45.7	43.7	46.4	679	47.5	46.4	48.5	45.5	49.6
638	45.1	44.3	45.8	43.7	46.5	680	47.6	46.5	48.6	45.6	49.6
639	45.1	44.4	45.8	43.8	46.5	681	47.6	46.5	48.7	45.6	49.7
640	45.2	44.4	45.9	43.8	46.6	682	47.7	46.6	48.8	45.6	49.8
641	45.3	44.5	46.0	43.9	46.7	683	47.8	46.6	48.8	45.6	49.9
642	45.3	44.6	46.0	43.9	46.8	684	47.8	46.6	48.8	45.6	49.9
643	45.4	44.6	46.1	44.0	46.9	685	47.9	46.7	48.9	45.7	49.9
644	45.4	44.7	46.2	44.0	46.9	686	47.9	46.7	49.0	45.7	50.0
645	45.5	44.7	46.2	44.1	47.0	687	48.0	46.8	49.0	45.8	50.1
646	45.6	44.8	46.3	44.1	47.1	688	48.0	46.8	49.1	45.8	50.2
647	45.6	44.9	46.4	44.2	47.2	689	48.1	46.9	49.2	45.9	50.3
648	45.7	44.9	46.5	44.2	47.2	690	48.1	46.9	49.3	45.9	50.4
649	45.7	45.0	46.5	44.3	47.3	691	48.2	46.9	49.3	46.0	50.5
650	45.8	45.0	46.6	44.3	47.4	692	48.3	46.9	49.3	46.0	50.6
651	45.9	45.0	46.6	44.3	47.4	693	48.3	47.0	49.4	46.1	50.7
652	45.9	45.1	46.7	44.4	47.5	694	48.4	47.1	49.5	46.1	50.7
653	46.0	45.2	46.8	44.4	47.6	695	48.4	47.1	49.5	46.1	50.8
654	46.1	45.2	46.9	44.5	47.7	696	48.5	47.1	49.6	46.2	50.9
655	46.1	45.3	46.9	44.5	47.8	697	48.5	47.2	49.7	46.2	51.0
656	46.2	45.3	47.0	44.6	47.9	698	48.6	47.2	49.7	46.3	51.1
657	46.2	45.4	47.1	44.6	47.9	699	48.7	47.3	49.8	46.3	51.2
658	46.3	45.4	47.2	44.7	48.0	700	48.7	47.3	49.8	46.4	51.2
659	46.4	45.5	47.2	44.7	48.1	701	48.8	47.3	49.9	46.4	51.3
660	46.4	45.5	47.3	44.8	48.2	702	48.8	47.4	50.0	46.4	51.4
661	46.5	45.6	47.3	44.7	48.2	703	48.9	47.4	50.0	46.4	51.5
662	46.5	45.6	47.4	44.8	48.3	704	48.9	47.4	50.1	46.5	51.5
663	46.6	45.7	47.5	44.8	48.3	705	49.0	47.5	50.2	46.5	51.6
664	46.6	45.7	47.5	44.8	48.3	706	49.0	47.6	50.3	46.5	51.7
665	46.7	45.8	47.6	44.8	48.4	707	49.1	47.7	50.4	46.6	51.8
666	46.8	45.8	47.7	44.9	48.5	708	49.2	47.7	50.5	46.6	51.9
667	46.8	45.9	47.7	44.9	48.6	709	49.2	47.8	50.5	46.7	51.9
668	46.9	45.9	47.8	45.0	48.6	710	49.3	47.8	50.6	46.7	52.0
669	46.9	45.9	47.8	45.0	48.7	711	49.3	47.8	50.7	46.7	52.1
670	47.0	46.0	47.9	45.1	48.8	712	49.4	47.9	50.7	46.7	52.2
671	47.1	46.1	48.0	45.1	48.9	713	49.4	48.0	50.8	46.8	52.3
672	47.1	46.1	48.1	45.1	49.0	714	49.5	48.0	50.9	46.9	52.4
673	47.2	46.1	48.1	45.3	49.1	715	49.6	48.1	51.0	46.9	52.4
674	47.2	46.2	48.2	45.3	49.2	716	49.6	48.1	51.0	46.9	52.5
675	47.3	46.2	48.2	45.3	49.2						
676	47.3	46.3	48.3	45.4	49.3						
677	47.4	46.3	48.4	45.4	49.4						

Technical
University of
Denmark

Frederiksborgsvej 399, Building 201
Risø Campus
Tlf. 4525 1700

www.fysik.dtu.dk