DTU Library

Testing the accuracy of a Bayesian central-dose model for single-grain OSL, using known-age samples

Guerin, Guillaume; Combès, Benoit; Lahaye, Christelle; Thomsen, Kristina Jørkov; Tribolo, Chantal; Urbanova, Petra; Guibert, Pierre; Mercier, Norbert; Valladas, Hélène

Testing the accuracy of a Bayesian central-dose model for single-grain OSL, using known-age samples

Guillaume Guérin[1*], Benoit Combès[1], Christelle Lahaye[1], Kristina J. Thomsen[2], Chantal Tribolo[1], Petra Urbanova, Pierre Guibert[1], Norbert Mercier[1], Hélène Valladas[3]

[1] *CNRS - Université Bordeaux Montaigne, UMR 5060, IRAMAT-CRP2A, Maison de l'archéologie, Esplanade des Antilles, 33607 Pessac cedex, France*
[2] *Center for Nuclear Technologies, Technical University of Denmark, DTU Risø Campus, DK-4000 Roskilde, Denmark.*
[3] *LSCE/IPSL, UMR CEA-CNRS-UVSQ, Avenue de la Terrasse, 91198 Gif sur Yvette Cedex, France.*
* Corresponding author: guillaume.guerin@u-bordeaux-montaigne.fr

Abstract

While reviews of comparisons between multi-grain OSL ages and independent chronological information are available in the literature, there is hardly any such performance test for single-grain OSL ages. And yet, this is all the more needed as the interpretation of single-grain dose distributions remains a difficult task, given the typically considerable dispersion in equivalent dose values measured by OSL – and the numerous sources of such dispersion in measurements. In this study, we present the OSL study of 19 samples for which independent age control is available, and whose ages range from 2 to 46 ka. Based on multi-grain OSL age estimates, these samples are presumed to have been both well-bleached at burial, and unaffected by mixing after deposition. Two ways of estimating single-grain ages are then compared: the standard approach on the one hand, consisting of applying the Central Age Model to D_e values determined with the Analyst software; on the other hand, the central dose model recently proposed by Combès et al. (Combès, B., Philippe, A., Lanos, P., Mercier, N., Tribolo, C., Guerin, G., Guibert, P., Lahaye, C., in press. Quaternary Geochronology). The median of the relative discrepancy between single-grain OSL and reference ages is about twice as large for the standard approach (12%) than with the Bayesian model (7%). Statistical tests show that, based on our (limited) data set, the difference between the two models seems to be significant for samples in the age range 4-46 ka. Finally, the influence of various factors on the (in-)accuracy of single grain OSL ages is discussed; it appears that the accuracy of ages estimated in a standard way decreases when age is increased, while the Bayesian model seems more robust. This study also shows that (i) there is no 20 % limit on the CAM overdispersion parameter for well-bleached samples; (ii) dose recovery experiments do not seem to be a very reliable tool to estimate the accuracy of a SAR measurement protocol for age determination.

Highlights

- Samples with independent age control are dated by OSL

- Their ages range from 2 to 46 ka

- Multi-grain OSL ages indicate no bleaching/mixing problem

- Single-grain OSL is analysed using two different models

- A Bayesian approach is expected to lead to improved accuracy

Keywords

Single-grain OSL; Accuracy tests; Bayesian statistics; Central Age Model; Comparative study

1

1. Introduction

Quartz OSL dating (Huntley *et al.*, 1985) has been increasingly applied in geochronology over the last 30 years. The development of Single Aliquot Regenerative (SAR: Murray and Wintle, 2000) dose measurement protocols has allowed determining individual equivalent dose (D_e) values for sets of grains usually called aliquots, whose size typically varies from several thousands of grains down to single-grains (Duller, 1999; Bøtter-Jensen *et al.*, 2000). Whereas OSL dating has been thoroughly tested against independent age control in a variety of environments for multi-grain aliquots (see reviews by Murray and Olley, 2002; Rittenour, 2008), only very few studies have focused on testing the accuracy of single-grain OSL age estimation in chronologically well-constrained case studies (Feathers, 2003; Olley *et al.*, 2004a; Arnold *et al.*, 2009; Olley *et al.*, 2004b; Demuro *et al.*, 2012). Moreover, in most cases the sedimentary deposition events studied concern the Holocene, *i.e.* "young" samples compared to the OSL age range (typically 100-200 ka, depending on various factors such as luminescence dose response of quartz, environmental dose rate). Validation studies of single-grain OSL as a reliable dating method are thus very scarce in the literature, especially for older samples.

The rationale for measuring single-grain OSL is to study sediment transport and deposition at the smallest possible scale, which can allow resolving issues such as post-depositional mixing of sedimentary units or insufficient resetting of OSL signal at the time of deposition. It has also been argued that the use of adequate grain selection criteria would allow removing 'junk-OSL' signals affecting the accuracy of age estimates (Jacobs *et al.*, 2013). In any case, while OSL dating has proved reliable for establishing Quaternary sediment deposition chronologies with multi-grain aliquots, it is not straightforward to assume that single-grain OSL ages will also be accurate. Indeed, while large multi-grain aliquots – due to important averaging effects – often display low scatter in D_e measurements, the situation becomes more complicated for single-grain D_e distributions, which typically exhibit considerable scatter '(and more specifically, overdispersion, i.e. dispersion that cannot be explained by analytical uncertainties but are due to several – presumed but hard to quantify – factors). Thus, even though the principles of OSL remain the same when the aliquot size is decreased, statistical analysis plays a fundamental role in deriving OSL ages from single-grain measurements. However, only very few studies have focused on this crucial step of age determination. In the literature, when no problem such as mixing or bleaching is suspected, only one age model is commonly used in OSL dating: the Central Age Model (CAM: Galbraith *et al.*, 1999). This model calculates a weighted geometric mean of individual equivalent doses; the overdispersion (OD) parameter is added in quadrature to measurement uncertainties derived from counting statistics, measurement reproducibility errors, and curve fitting uncertainties, to weight individual D_e values in the mean calculation.

The CAM requires, as input, lists of paired D_e values and uncertainties. As a consequence, the standard approach to OSL data analysis consists of estimating individual D_e values parameterised by Gaussian probability densities – typically obtained using the Analyst programme (Duller, 2007) or the R Luminescence package (Kreutzer *et al.*, 2012). Such parameterization can be problematic, because density probabilities obtained by the projection of luminescence measurements onto non-linear dose response curves are not symmetric D_e probability distributions. At present, Analyst allows calculating asymmetric errors on D_e estimates; however, the CAM is not suited to such data. This is one of the reasons that led Combès *et al.* (in press) to design a new luminescence central dose model using Bayesian statistics. In this model, it is assumed that the analysed sample is characterised by a central dose around which individual D_e values are distributed following a Cauchy distribution. Priors are defined for dose response curves, and posterior density probabilities allow estimating central doses that are expected to be more coherent with the experimental data.

In this study, we selected 19 presumably well-bleached samples associated with independent age control, mainly radiocarbon ages (Table 1; see also supplementary information). Expected OSL ages range from ~2 to 46 ka and come from a variety of environments. The purpose of this study is twofold: firstly, we

apply the CAM to D_e values obtained with Analyst, to derive "standard" single-grain OSL ages and estimate their accuracy, assuming that the reference age is the target of the OSL analysis. Secondly, the Bayesian model of Combès *et al.* (in press) is applied to the same samples; results are compared with both the standard OSL ages and the age references.

2. Samples and methods

2.1 Samples

All samples investigated here are the subject of separate, independent chronological studies that are either already published or in the course of publication. Therefore, only the most basic data will be given here. Table 1 lists all the studied samples, which mostly come from the archaeological record: samples Bdx 16045-16049 are mortar samples that were sampled from roman foundations of the medieval castle located in the city of Antibes (France; Urbanova, in prep.) and are supposedly the youngest samples of the series (~2 ka, based on the production age of ceramics found in a foundation trench of the castle, corroborated by archaeo-magnetism measurements). The inter-comparison sample comes from a beach-ridge located in Skagen (Denmark) and it has been selected as it has been the basis for a wide, luminescence community-scale study involving a number of OSL laboratories across the world (Buylaert *et al.*, 2006; Murray *et al.*, 2015; Guérin *et al.*, 2015a). We consider it as a reference sample, whose expected age (~4 ka) is taken as the average of all measurements (n=24, mainly on multi-grain aliquots) reported in the frame of the intercomparison study. Fluvial samples BR-2011-8, -11, -15 and -32 come from the Vale da Pedra Furada (Brazil), which is an important prehistoric site for the study of human dispersal on the American continent (Lahaye *et al.*, 2013; Boëda *et al.*, 2014; for the single-grain study, see Lahaye *et al.*, submitted); their expected ages range from 7 to 24 ka, based on the association with radiocarbon samples. Samples EBC 1 and 2 come from the Middle/Late Stone Age layers from Eland's Bay Cave (South Africa) and their expected ages are 23 and 37 ka, respectively (^{14}C ages; Tribolo *et al.*, in prep. a). Samples GDB 3 and 5 come from Goda Buticha, a cave site in Ethiopia (Tribolo *et al.*, in prep. b); their expected ages are 43 and 7 ka, respectively (^{14}C ages). Samples 092201, -02, -03 and -04 come from Middle and Late Palaeolithic layers from the Bordes-Fittes rock shelter (France), and the expected ages range from 23 to 45 ka (^{14}C ages; Aubry *et al.*, 2012; 2014). Finally sample FER 3 is a colluvial sample from the Late and Middle Palaeolithic La Ferrassie rock shelter (France), with an expected age of 46 ka (^{14}C age; Guérin *et al.*, 2015b).

For a number of the selected samples, more than one associated radiocarbon age is available (Table S1). Here, to make comparisons easier, we have decided to calculate, for each OSL sample, a unique average reference age. To do so, we have first calibrated the radiocarbon ages with IntCal13 (Reimer *et al.*, 2013). Then, the obtained 95% credible intervals (CI) have been combined as follows: for each OSL sample, the lowest and highest age boundaries obtained with radiocarbon are taken as the limits of the 'average' reference 95 % CI. In other words, for each OSL sample the average reference age is calculated as x ± y, where x is the average between the lowest and highest individual 95 % CI boundaries, and y is taken as the quarter of the difference between these two extreme ages. Hence, the interval for the average reference age covers all individual 95 % CIs.

Coarse quartz grains (generally 180-250 µm, sometimes 200-250 µm) were extracted from all samples using standard laboratory procedures, under controlled dim light conditions. Single-grain OSL measurements were all performed on automated Risø TL/OSL DA-20 readers, in two different laboratories : at Risø-DTU (Roskilde, Denmark) for samples 092201, -02, -03 and -04, the intercomparison sample and sample FER 3; at IRAMAT-CRP2A (Bordeaux, France) for the remaining samples. Five different users are involved in the present study (PU, CT, CL, KJT and GG). Measurement conditions and SAR parameters are summarised in Table S2.

For all samples, K, U, Th contents were determined using high resolution gamma spectrometry and then converted in beta dose rates using dose-rate conversion factors from Guérin *et al.* (2011), and beta attenuation factors from Guérin *et al.* (2012). In some cases these radioelements contents were also converted in gamma dose rates (samples 092201, -02, 03, -04, and the intercomparison sample); for the rest of the samples, gamma dose rates were measured in situ – either using a portable gamma spectrometer (with the so-called threshold technique, following Mercier and Falguères, 2007; Miallier *et al.*, 2009; Guérin and Mercier, 2011: samples BR-2011-8, -11, -15 and -32; GDB 3 and 5) or Al_2O_3:C dosimeters buried in the sediments (following the procedure described by Richter *et al.*, 2010: samples FER 3, Bdx 16041-45, EBC 1 and 2).

The list of studied samples thus come from a variety of geomorphological and archaeological settings, have been measured in two different laboratories by various users, their age span a good part of the radiocarbon dating age range (from 2 to 46 ka), and different techniques have been used to measure gamma dose rates: as a result, we believe that this corpus is – to some degree at least – representative of typical OSL studies aiming at establishing Quaternary geochronology.

2.2 Multi-grain OSL measurements

For 13 out of the 19 samples, multi-grain OSL ages were obtained; depending on the authors responsible for the different studies, either the CAM or a simple arithmetic average of measured D_e values has been used to derive ages. It is not our aim to discuss here the merits of either approach on multi-grain D_e values; however, because of generally low scatter in D_e estimates, ages using the CAM or a plain average should not be very different. These ages are plotted as a function of reference ages in Fig. 1a and show a good correlation. The median ratio of OSL to reference age is 0.969, and the median relative discrepancy between OSL and reference ages is 8%; this figure is comparable with the size of OSL age uncertainties (Guérin *et al.*, 2013, argue that luminescence ages cannot have realistic age uncertainties of less than ~5%), and so this comparison suggests that all samples measured in multi-grain aliquots were most likely well-bleached at deposition, and unaffected by significant post-depositional mixing. It should be noted here that, given the (low) number of samples, we consider the median to be more representative of the central tendency than the average age ratio (note the presence of outliers, see Table 1).

Obviously, the OSL to reference age ratios cannot be directly used as indicator of accuracy, in the sense that the true sediment deposition age remains unknown. In particular, the reference ages (mainly obtained with radiocarbon) do not date the same event as OSL. However, in the following we assume that the difference between the ages targeted by OSL and ^{14}C is negligible compared to the age uncertainties (in particular of OSL ages, for which the uncertainties are greater than 5%); the very good agreement between multi-grain OSL ages and reference ages tends to strengthen this view.

2.3 The analysis models

Two analysis models are compared in this study. The first is the most commonly employed analysis model, where the CAM is applied to individual D_e values calculated with Analyst (or a similar software, such as the Luminescence R package) – hereafter this approach will be referred to as the standard approach. In this approach, luminescence data and regenerative doses are combined to fit a curve to individual dose response curves; the projection of the natural corrected OSL signal on this curve leads to individual D_e estimates.

On the other hand, Combès *et al.* (in press) recently proposed a new model to analyse OSL data in a Bayesian framework. In this approach, no inference is made on individual aliquots (*e.g.,* individual equivalent doses are not parameterised independently, contrary to what is obtained with more standard analysis tools such as Analyst or the R luminescence package): selected sets of grains are simultaneously analysed, assuming that they belong to a population characterized by a central age distribution. Then, while

in the CAM individual D_e values are assumed to be distributed following a lognormal distribution, in the Bayesian model the dispersion of D_e values is described by a Cauchy distribution – which provides better robustness properties (Combès *et al.*, in press). Because the inference on the central age is based on all measured luminescence data and associated regenerative doses, rather than drawn from individual D_e values, and because no assumption is made regarding the shape of these individual (*i.e.*, at the aliquot level) D_e probability distributions, Combès *et al.* (in press) argue that their Bayesian central dose model provides results that are more coherent with the raw measurement data. As a consequence, this could lead to more accurate results than those obtained using the standard approach.

In both approaches, users have to select which grains to include in the analysis. The purpose of this study is not to discuss the relevance of the various selection criteria stated in the literature. As a consequence, we have decided to follow the different users' choices (see Table S2); the same sets of grains have been analysed with both models. Similarly, the choice of signal and background summation intervals vary from one author to the other; these parameters can also be found in the original publications – the important point here is that the same signal summation intervals have been used in both analyses.

Another user-dependent choice is the nature of functions that are used for fitting dose response functions. Combès *et al.* (in press) chose to implement their model with exponential + linear functions; for the purpose of comparison with the standard approach, in Analyst the same curve fitting option has been used (it should be noted here that, depending on users, different fitting choices have been made in the original publications regarding the samples studied here; such fitting discrepancies explain differences between results in this paper and those, for the same samples, from the original publications). However, even though the functions used are the same, the domain on which the coefficient of the linear component is defined is not the same in both approaches: in the Bayesian model, this coefficient is forced to be positive, while in Analyst it can be negative. As a result, in the Bayesian approach no grain can be in saturation, since the OSL signal grows indefinitely, while for some grains Analyst does not give a D_e value (for example, when the natural OSL signal does not intercept the dose response curve, or when two solutions exist). It may also happen that the uncertainty given by Analyst is 'infinite'. In practice, it means that in the standard approach, some grains will not be considered after the Analyst step, *i.e.* no dose value will be included in the CAM calculations. Finally, in the Bayesian approach the dose response curve is not forced through the origin; for consistency, we have systematically un-ticked the box 'Force growth curve through origin' in Analyst.

3. Results: compared single-grain ages

In the single grain OSL literature, the choice of age models is generally discussed in the light of various factors, among which the deposition process of the sediments, potential post-depositional disturbances, but also in the light of the overdispersion (OD) value determined for the measured D_e distributions using the CAM. In particular, it is often argued in the literature that typical OD values for well-bleached, single dose component samples are lower than 20 %. For example, Jacobs *et al.* (2008) state that, based on the work by Olley *et al.* (2004b) "samples consisting of well-bleached quartz grains commonly have De distributions that are overdispersed by up to 20%"; then, they "ran the FMM using σ_d values of between 10 and 20%" (Jacobs *et al.*, 2008; Supporting Online Material, p. 13). On the same subject of the application of the FMM to identify well-bleached components, and in particular when justifying that a maximum of 20 % for the OD value is used to run the FMM, Jacobs *et al.* (2011) state that "there is little merit […] in using very high OD values *(NB: greater than 20%)*, as the latter are poor analogues for a well-bleached sample" (see Supplementary Material, p. 17). This supposed 20 % limit was originally based on a study comprising 7 samples (Olley *et al.*, 2004b), of which only 2 gave an overdispersion value greater than 20 %; of these two, only one (sample OSL 6) clearly required the minimum age model to be applied for a

better agreement with radiocarbon (for sample OSL 5, both the MAM and the CAM give results consistent with radiocarbon data, and it is hard to tell if one of the two models is better than the other).

On the other hand, Mayya *et al.* (2006) and Guérin *et al.* (2015a) have shown that dose rate variations at the mm-scale, for example due to the presence of potassium feldspar grains, can lead to continuous D_e distributions with arbitrarily large OD values. Fig. 2 shows the histogram of OD values obtained for the dataset used in this study; no sample exhibits an OD value lower than 20 %, and the OD values range from 21 ± 1 % to 62 ± 6 %. The good agreement between multi-grain quartz OSL ages and reference ages argues against bleaching and post-depositional mixing; we think that this is convincing evidence that there is no 20 % limit (or actually any other such limit) on the OD value for well-bleached samples – as already stated by Thomsen *et al.* (2012; their Fig. 1 and associated discussion).

In Fig. 1b, single-grain ages obtained with the Bayesian model and with the standard approach are plotted as a function of associated reference ages, and show overall good consistency (see also Table 1). Linear regression of the two data sets seems to indicate that the Bayesian ages are closer to the references than the standard single-grain ages (see also Table 1, where the ratio of these single-grain OSL to reference ages is given). The median – as stated earlier, due to the limited number of samples we use the median of this ratio to estimate the central tendency – ratio is 0.927 for the Bayesian model, compared to 0.879 for the standard approach.

Thus far, we have taken the good agreement between multi-grain OSL and reference ages as an argument for the absence of bleaching and mixing problems of the studied samples. However, multi-grain OSL age estimates are not available for all samples; if we consider only those samples for which multi-grain OSL ages are available (n=13), the median ratio of OSL to reference age is 0.927 with the Bayesian model, and 0.866 for the standard approach. For the remaining samples, these ratios are 0.933 and 1.001 (n=6), respectively. In the latter set of samples, since the agreement between single-grain OSL and reference ages is better than in the former (presumed to be well-bleached, based on multi-grain OSL ages), we have to conclude that these samples are also unaffected by bleaching and mixing problems. In such conditions, the median ratios obtained with the two tested models seem to confirm that the ages obtained with the Bayesian central dose model are more accurate that the ages obtained in the standard manner. Nevertheless, both analysis models systematically underestimate the reference ages, by a greater amount than multi-grain OSL ages (see section 2.2).

In Fig. 3, the relative difference between the OSL ages calculated with the standard approach, and the reference ages is plotted as a function of the same quantity when the OSL age calculation is performed with the Bayesian central dose model. Points lying on top of the 1:1 line (12 out of 19) indicate a better accuracy of the Bayesian central dose model compared to the CAM applied to D_e values determined using Analyst. Furthermore, it appears that 4 of the 7 points lying below the 1:1 line belong to the youngest sample set (Bdx 16046 to 16049).

4. Discussion

In this comparison between OSL and reference ages (mainly from radiocarbon), three different ways of obtaining OSL age estimates have been tested. It appears that multi-grain OSL ages are the closest to the reference ages, followed by single-grain ages calculated using the Bayesian model of Combès *et al.* (in press) and finally by single-grain ages using the standard approach (Analyst and CAM). But how significant is the difference observed between Bayesian and standard single grain OSL ages? To answer this question, we performed a paired Wilcoxon-Mann-Whitney test with the alternative hypothesis (H_1): "median (E_{bayes}) ≠ median ($E_{Standard}$)" (where E_{Bayes} and $E_{Standard}$ are respectively the random variables associated with the distribution of errors of the Bayesian and the standard method). This yields a p-value of 0.28 that fails to reject the null hypothesis (H_0) at the 0.05 confidence level. Similarly, we performed a sign

test (Pesarin and Salmaso, 2010) with the alternative hypothesis (H_1): "$Pr(E_{bayes} < E_{standard}) \neq 0.5$" which gives a similar result (p-value of 0.22). It should be noted that the first test is more informative, but requires E_{bayes} - $E_{Standard}$ to have a symmetric distribution. Conversely, the sign test does not require such a hypothesis. Thus, with our limited data set it appears that the difference between the two analysis models is not significant. Further tests with more samples will thus be needed to assess whether or not the differences that we see in our data set are real. Nevertheless, the purpose of the following discussion is to discuss various factors – error properties, age, overdispersion, etc. – that could influence the accuracy of the tested age models. But first, the use of the word 'accuracy' may need to be discussed, in particular when comparing OSL with radiocarbon ages beyond 40 ka.

How reliable are the reference ages?

In this study, the radiocarbon ages included for comparison with OSL ages go up to ~46 ka and are thus close to the limit of application of the method. How much confidence can we have in such data? First, since the radiocarbon community now works with a calibration curve up to 50 ka, we see no *a priori* reason to reject data (see for example the discussion in Talamo *et al.*, 2012). This being said, for very old samples even the tiniest contamination may have important consequences on the ages. So when approaching the limits of radiocarbon we would expect radiocarbon ages younger than OSL ones (*e.g.*, Long and Shen, 2015). If we had a systematic underestimation of radiocarbon compared to OSL, we would have to discuss this problem in detail – and perhaps eventually dismiss the ^{14}C ages. But in the case of the present study, for all of the old samples we observe radiocarbon ages greater than OSL ages (Table 1). So we assume that our radiocarbon ages are not problematic; if they were, it would only mean that OSL ages are even more problematic than we think.

Dose recovery tests and accuracy of associated ages

It is common practice in luminescence dating to present quality-assurance tests, such as results of dose recovery tests. The aim of such tests is to check whether the SAR measurement protocol leads to accurate determination of laboratory doses given after OSL signal resetting by a solar lamp or by optical stimulation in a luminescence reader. The results of dose recovery tests are presented in supplementary data (Table S3). The median absolute departure from unity is 3.4 % with the Bayesian model, and 4.5 % for the standard approach (Table S3). Thus, it seems that for these dose recovery tests, the measured to given dose ratios estimated with the Bayesian model are more accurate than those estimated with the standard approach (this confirms results of Combès *et al.*, in press). Fig. S1 presents normalised ages (i.e., the ratio of OSL to reference age) as a function of dose recovery ratios. For both analysis models, no significant correlation between these two quantities is observed, and the slope of the lines fitted to the data is not statistically different from zero (0.05 level). Thus, either the results of these tests are not representative of the measurement of natural doses; or the apparent inaccuracy of single-grain ages is mainly driven by dose rate errors. But because (i) multi-grain OSL ages have been shown to be accurate in reviews by Murray and Olley (2002) and Rittenour (2008), and (ii) in the present study multi-grain OSL ages are in good agreement with the reference ages, we argue that the systematic underestimation affecting single-grain age estimates must be mainly due to single-grain D_e measurements and analyses. From this, we conclude that dose recovery ratios – at least for the samples studied here – are not a good indicator of the accuracy of ages obtained with the SAR protocol (as already suggested by Murray and Wintle, 2003).

Factors influencing the accuracy of single grain OSL ages

Guérin *et al.* (2015a) recently argued that the CAM, because it calculates a geometric mean of D_e values, might underestimate OSL ages when the OD is 'large' (the bigger the dispersion in D_e values, the greater the discrepancy between geometric and arithmetic means; since dose rate measurements provide arithmetic means, this could lead to age underestimation). We have tested this hypothesis by plotting the

relative difference between OSL ages calculated in the standard manner and reference ages, as a function of CAM OD; no correlation appears (data not shown here). In other words, the aforementioned hypothesis of Guérin *et al.* (2015a) is not verified by the studied dataset – the age underestimation with the standard approach does not seem to be related to large OD values.

One assumption underlying the Central Age model is that D_e uncertainties are driven by multiplicative error properties (*i.e.*, the size of errors is proportional to dose – see Galbraith and Roberts, 2012). To test the validity of this assumption, we looked at two samples for which the Bayesian model is very close to their reference age, while the standard approach gives a significant age underestimate (samples 092201 and 092202). In Figs. S2a and S2b, relative uncertainties are plotted as a function of dose. No trend is observable, *i.e.* relative errors seem to be independent of dose estimates; hence, the bias observed for single-grain standard age estimates cannot be attributed to the discordance between error properties and the analysis model (CAM – at least for these two samples).

Another factor that could explain the discrepancy between the standard and the Bayesian approaches is increasing dose. Fig. 4a shows the relative difference between these two analytical approaches as a function of CAM dose. Despite scattered data, there is a correlation: the relative difference between the two models is increased when the dose is increased; more precisely, as dose is increased the Bayesian model tends to give greater age estimates than the standard approach. Another way to look at this is to observe the Bayesian dose as a function of CAM dose (Fig. 4b). In other words for constant dose rates, the older the age of a sample, the bigger the difference between the ages obtained with the two approaches will be. At this stage, it is difficult to explain this difference with any degree of certainty: however, two reasons might lead to greater ages using the Bayesian model, compared to the standard approach. The first reason is curve fitting: as explained above (section 2.3), in the Bayesian model (but not in Analyst) OSL signals are forced to increase monotonically with dose; in Analyst, the exponential + linear fitting option allows normalised OSL to decrease with dose, in which case multiple solutions exist for projecting the natural OSL on the dose response curve. In such conditions, the considered grain has to be discarded (whereas it remains included in the Bayesian analysis). This difference in curve fitting may explain – at least partly – why the Bayesian model allows exploring higher dose values. A second potential explanation is that in the standard approach, D_e values are parameterised by Gaussian probability distributions; the central value of this distribution corresponds to the projection of the natural normalised OSL signal on the Dose Response Curves (DRC). Then, uncertainties are calculated from the length of the dose interval obtained by projecting the 68% confidence interval of normalised OSL signals onto the dose response curve (Duller, 2007). In cases where the dose response is not linear (and more specifically sub-linear, close to saturation), this will lead to underestimation of the dose – by contrast, the Bayesian model does not suffer from such limitations (it should be emphasized here that Analyst allows calculating asymmetric uncertainties, but such data cannot be fitted by the CAM).

In the next step, we investigated this trend by comparing the single-grain OSL ages with the average reference ages. In Fig. 5, the ratio of OSL (obtained with the Bayesian model, Fig. 5a; with the standard approach, Fig. 5b) to average reference age is plotted as a function of reference age. It appears that the accuracy of the ages calculated using the Bayesian central dose model does not significantly vary with increasing age, while ages calculated in the usual manner seem to become less accurate when the age is increased. Indeed, in the latter case the slope of the fitted line (dashed line in Fig. 5b) is significantly different from zero (-2.5 ± 1.0 10^{-3} ka^{-1}; for the Bayesian approach, the slope is 0.2 ± 1.1 10^{-3} ka^{-1}). This observation, together with the fact that for four out of the five 2 ka old mortar samples seem to get better agreement with the reference age using the standard approach than using the Bayesian model, led us to try the two statistical tests afore-mentioned (Wilcoxon-Mann-Whitney and sign test) after excluding this batch of five mortars coming from the same building. Both tests yield to reject the null hypothesis (H_0) at the 0.05 confidence level (p-value of 0.02) and suggest a significant difference between the two analysis models.

Obviously, such results should be seen with great caution; we take them only to indicate that age may be an important factor in resolving the apparently decreasing accuracy of the standard approach for calculating single grain OSL ages.

The potential influence of D_e distribution models

Finally, it should be stressed that the two single-grain analysis models tested in this study are conceptually different: on the one hand, one considers a Bayesian inference on a hierarchical model; on the other hand, a sequence of error propagations followed by a frequentist confidence interval estimation (standard approach). Another important difference between the two models consists of the central D_e model: a Cauchy distribution in the Bayesian approach, compared to a lognormal distribution in the standard approach. In our view, there is little doubt that the inference drawn directly from the measurement dataset rather than from a series of intermediate parameterisation steps provides a more satisfying treatment of data. This being said, a significant part of the difference between the results obtained with the two models is likely to come from the distributions describing the dispersion of individual D_e values around the dose of interest. As pointed out by Combès *et al.* (in press), Cauchy distributions are heavy-tailed and, as such, significantly less sensitive to contaminant grains than lognormal distributions. This confers to the Cauchy based model great robustness properties and this is what led the choice of Combès *et al.* (in press), since single-grain D_e distributions often contain outliers. However, in cases of strongly asymmetric distributions of individual D_e values around the characteristic dose, the robustness property of the Cauchy distribution can be counter-productive and result in poorer accuracy (this seems to be the case for the mortar sample set, in particular for samples Bdx 16046-49). From the present study, it is not clear which equivalent dose model is the optimal one as one cannot distinguish the merits of the approach (standard, stepwise compared to Bayesian, global) from those of the equivalent dose model (Cauchy compared to lognormal distributions). Answering this question goes beyond the scope of this study but would be a promising extension of this work.

5. Conclusion

This study provides a review of comparisons between single-grain OSL ages and independent chronological information. A set of 19 samples have been analysed; OSL single-grain ages obtained with two different approaches have been compared with multi-grain OSL and reference ages (mainly radiocarbon ages, ranging from 2 to 46 ka). Firstly, multi-grain OSL ages appear to be the closest to the reference ages.

Most importantly, the Bayesian central dose model proposed by Combès *et al.* (in press) seems to give more accurate single-grains OSL ages than the standard approach combining Analyst and CAM: the median of the relative discrepancy between single-grain OSL and reference ages is about twice as large for the standard approach (12%) than for the Bayesian model (7%). The difference between the two models, estimated with all the samples tested (n=19), is not statistically significant; however, it becomes significant when excluding the batch of the five youngest mortar samples. Interestingly, our dataset suggests that the accuracy of single-grain ages estimated with the standard approach may decrease when the age is increased; by contrast, the Bayesian model of Combès et al. (in press) appears to be a promising alternative for single-grain OSL analysis (the difference between the models seems to become significant when age is increased). Further testing of the model will be required to assess if the differences between the analysis models are significant, and if so for what reasons.

General considerations regarding the analysis of single-grain OSL data can also be drawn from this study: (i) there is certainly no 20 % limit on the CAM overdispersion parameter for well-bleached samples (it should be emphasized that the possibility of post-depositional mixing and/or insufficient OSL signal resetting at deposition can reasonably be discarded from our study based on the agreement between OSL

multi-grain and reference ages); this corroborates studies from a dose rate perspective. (ii) Based on our data set, dose recovery experiments do not seem to be a very reliable tool to estimate the accuracy of a SAR measurement protocol for age determination.

Acknowledgements

References

Aubry, T., Dimuccio, L.A., Almeida, M., Buylaert, J.P., Fontana, L., Higham, T., Neves, M.J., Peyrouse, J.B., Walter, B., 2012. Stratigraphic and technological evidences from the Middle Palaeolithic-Châtelperronian-Aurignacian record at the Bordes-Fitte rockshelter (Roches d'Abilly site, Central France). Journal of Human Evolution 62, 116-137.

Aubry, T., Dimuccio, L.A., Buylaert, J.P., Liard, M., Murray, A.S., Thomsen, K.J., Walter, B., 2014. Middle-to-Upper Palaeolithic site formation processes at the Bordes-Fitte rockshelter (Central France), Journal of Archaeological Science 52, 436-457.

Boëda E., Clemente-Conte I., Fontugne M., Lahaye C., Pino M., Felice G.D., Guidon N., Hoeltz S., Lourdeau A., Pagli M., Pessis A.M., Viana S., Da Costa A., Douville E., 2014. A new late Pleistocene archaeological sequence in South America: the Vale da Pedra Furada (Piauí, Brazil). Antiquity, 88, 927-941.

Bronk Ramsey, C., 2009. Bayesian analysis of radiocarbon dates. Radiocarbon, 51, 337-360.

Buylaert, J.-P., Ankjaergaard, C., Murray, A.S., Nielsen, A., 2006. A proposed laboratory intercomparison sample based on a beach-ridge sand from Skagen (Denmark), poster. In: UK-LED Meeting, Liverpool.

Bøtter-Jensen, L., Bulur, E., Duller, G.A.T., Murray, A.S., 2000. Advances in luminescence instrument systems. Radiation Measurements 32, 523–528.

Combès, B., Philippe, A., Lanos, P., Mercier, N., Tribolo, C., Guerin, G., Guibert, P., Lahaye, C., in press. A Bayesian central equivalent dose model for optically stimulated luminescence dating. Quaternary Geochronology.

Demuro, M., Froese, D.G., Arnold, L.J., Roberts, R.G., 2012. Single-grain OSL dating of glaciofluvial quartz constrains Reid glaciation in NW Canada to MIS 6. Quaternary Research 77, 305–316.

Duller, G.A.T., 2007. Analyst. Manual, 1–45.

Duller, G.A.T., Bøtter-Jensen, L., Murray, A.S., Truscott, A.J., 1999. Single grain laser luminescence (SGLL) measurements using a novel automated reader. Nuclear Instruments and Methods B, 155, 506–514.

Feathers, J.K., 2003. Single-grain OSL dating of sediments from the Southern High Plains, USA. Quaternary Science Reviews 22, 1035–1042.

Galbraith R.F., Roberts R.G., Laslett G.M., Yoshida H., 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, 339–364.

Galbraith, R.F., Roberts, R.G., 2012. Statistical aspects of equivalent dose and error calculation and display in OSL dating: an overview and some recommendations. Quaternary Geochronology 11, 1-27.

Guérin G., Mercier, N., 2011. Determining gamma dose rates by field gamma spectroscopy in sedimentary media: results of Monte Carlo simulations. Radiation Measurements 46, 190-195.

Guérin, G., Mercier, N. and Adamiec, G, 2011. Dose-rate conversion factors: update. Ancient TL, 29 (1), 5-8.

Guérin, G., Mercier, N., Nathan R., Adamiec, G., Lefrais, Y., 2012. On the use of the infinite matrix assumption and associated concepts: a critical review. Radiation Measurements, 47, 778-785.

Guérin, G., Murray A. S., Jain M., Thomsen K. J., Mercier, N., 2013. How confident are we in the chronology of the transition between Howieson's Poort and Still Bay? Journal of Human Evolution 64, 314-317.

Guérin, G. Jain, M., Thomsen, K. J., Murray, A.S., Mercier, N., 2015a. Modelling dose rate to single grains of quartz in well-sorted sand samples: the dispersion arising from the presence of potassium feldspars and implications for single grain OSL dating. Quaternary Geochronology 27, 52-65.

Guérin, G., Frouin, M., Talamo, S., Aldeias, V., Bruxelles, L., Chiotti, L., Dibble, H. L., Goldberg, P., Hublin, J.-J., Jain, M., Lahaye, C., Madelaine, S., Maureille, B, McPherron, S. P., Mercier, N., Murray, A. S., Sandgathe, D., Steele, T. E., Thomsen, K. J., Turq, A., 2015b. A Multi-method Luminescence Dating of the Palaeolithic Sequence of La Ferrassie Based on New Excavations Adjacent to the La Ferrassie 1 and 2 Skeletons. Journal of Archaeological Science, in press.

Huntley, D.J., Godfrey-Smith, D.I., Thewalt, M.L.W., 1985. Optical Dating of sediments. Nature 313, 105-107.

Jacobs, Z., Roberts, R.G., Galbraith, R.F., Deacon, H.J., Grün, R., Mackay, A., Mitchell, P., Vogelsang, R., Wadley, L., 2008. Ages for the Middle Stone Age of southern Africa: implications for human behavior and dispersal. Science 322, 733-735.

Jacobs, Z., Meyer, M.C., Roberts, R.G., Aldeais, V., Dibble, H., El Hajraoui, M.A., 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, 3631-3643.

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

Lahaye, C., Hernandez, M., Boëda, E., Felice, G. S., Guidon, N., Hoeltz, S., Lourdeau, A., Pagli, M., Pessis, A.-M., Rasse, M., Viana, S., 2013. Human occupation in South America by 20,000 BC: the Toca da Tira Peia site, Piauí, Brazil. Journal of Archaeological Science 40, 2840-2847.

Lahaye, C., Guérin, G., Boëda, E., Fontugne, M., Frouin, M., Clemente-Conte, I., Pino, M., Felice, G.D., Guidon, N., Lourdeau, A., Pagli, M., Pessis, AM., Da Costa, A. New insights into a late-Pleistocene human occupation in America: the Vale da Pedra Furada complete chronological study. Submitted to Quaternary Geochronology.

Long, H., Shen, J., 2014. Underestimated 14C-based chronology of late Pleistocene high lake-level events over the Tibetan Plateau and adjacent areas: evidence from the Qaidam Basin and Tengger Desert. Science China: Earth Sciences 58, 183-194.

Mayya, Y.S., Morthekai, P., Murari, M.K., Singhvi, A.K., 2006. Towards quantifying beta microdosimetric effects in single-grain quartz dose distribution. Radiation Measurements, 41, 1032-1039.

Mercier, N., Falguères, C. 2007. Field gamma dose-rate measurement with a NaI(Tl) detector : re-evaluation of the "threshold" technique. Ancient TL, 1-4.

Miallier, D., Guérin, G., Mercier, N., Pilleyre, T. and Sanzelle, S., 2009. The Clermont radiometric reference rocks: a convenient tool for dosimetric purposes. Ancient TL, 27, 37-44.

Murray, A. S., Wintle, A. G., 2000. Luminescence dating of quartz using an improved single-aliquot regenerative-dose protocol. Radiation Measurements 32, 57-73.

Murray, A. S., Wintle, A. G., 2003. The single aliquot regenerative dose protocol: potential for improvements in reliability. Radiation Measurements 37, 377-381.

Murray, A.S., Olley, J.M., 2002. Precision and Accuracy in the Optically Stimulated Luminescence Dating of Sedimentary Quartz: A Status Review. Geochronometria 21, 1-16.

Murray, A. S., Buylaert, J.-P., Thiel, C., 2015. A Luminescence Dating Intercomparison Based on a Danish Beach-ridge Sand. Radiation Measurements, in press.

Olley, J.M., Pietsch, T., Roberts, R.G., 2004a. Optical dating of Holocene sediments from a variety of geomorphic settings using single grains of quartz. Geomorphology 60, 337-358.

Olley, J.M., Deckker, P.D., Roberts, R.G., Fifield, L.K., Yoshida, H., Hancock, G., 2004b. Optical dating of deep-sea sediments using single grains of quartz: a comparison with radiocarbon. Sedimentary Geology 169, 175-189.

Pesarin, F., Salmaso, L., 2010. Permutation tests for complex data: Theory, applications and software. New York: Wiley.

Reimer, P. J., Bard, E., Bayliss, A., Beck, J. W., Blackwell, P. G., Bronk Ramsey, C., Grootes, P. M., Guilderson, T. P., Haflidason, H., Hajdas, I., HattŽ, C., Heaton, T. J., Hoffmann, D. L., Hogg, A. G., Hughen, K. A., Kaiser, K. F., Kromer, B., Manning, S. W., Niu, M., Reimer, R. W., Richards, D. A., Scott, E. M., Southon, J. R., Staff, R. A., Turney, C. S. M., & van der Plicht, J., 2013. IntCal13 and Marine13 Radiocarbon Age Calibration Curves 0-50,000 Years cal BP. Radiocarbon, 55, 1869-1887.

Richter, D., Dombrowski, H., Neumaier, S., Guibert, P., Zink, A., 2010. Environmental gamma dosimetry for in-situ sediment measurements by OSL of a-Al2O3:C. Radiation Protection Dosimetry 141, 27-35.

Rittenour, T.M., 2008. Luminescence dating of fluvial deposits: applications to geomorphic, palaeoseismic and archaeological research. Boreas, 37, 613-635.

Talamo, S., Hughen, K. A., Kromer, B. Reimer, P. J., 2012. Debates over Palaeolithic chronology – the reliability of 14C is confirmed. Journal of Archaeological Science 39, 2464-2467.

Thomsen K. J., Murray A. S., Jain M., 2012. The dose dependency of the over-dispersion of quartz OSL single grain dose distributions. Radiation Measurements, 47, 732-739.

Tribolo C, Mercier N, Valladas H., Miller C., Mentzer S., Parkington J., Porraz G. (in prep., a): Luminescence dating at Elands Bay Cave. Southern African Humanities

Tribolo C., Asrat A., Bahain J.-J., Chapon C., Douville E., Fragnol C., Hernandez M., Hovers E., Leplongeon A., Pleurdeau D., Pearson O., Puaud S., Assefa Z. (in prep, b). Geochronological and sedimentological analyses of the Later and Middle Stone Sequence of Goda Buticha (Ethiopia).

Urbanova, P., in preparation. Researches on direct dating of constructions: Investigations of the potential of optically stimulated luminescence to date archaeological mortars. PhD thesis, University Bordeaux Montaigne.
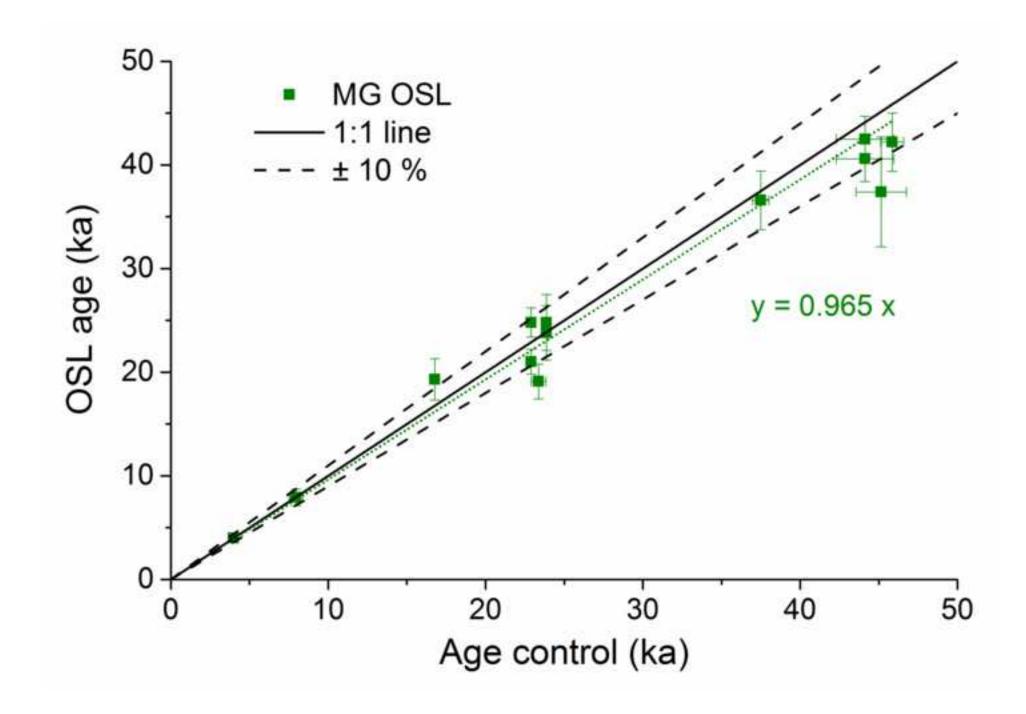
Figure captions

Fig. 1. OSL ages as a function of reference ages: (a) multi-grain OSL ages (calculated using the standard approach); (b) single-grain OSL ages. Blue triangles represent ages calculated with the standard approach (Analyst and CAM), while red circles represent ages calculated with the Bayesian central dose model of Combès *et al.* (in press). Inset in (b): same as (b), but focusing on younger samples.
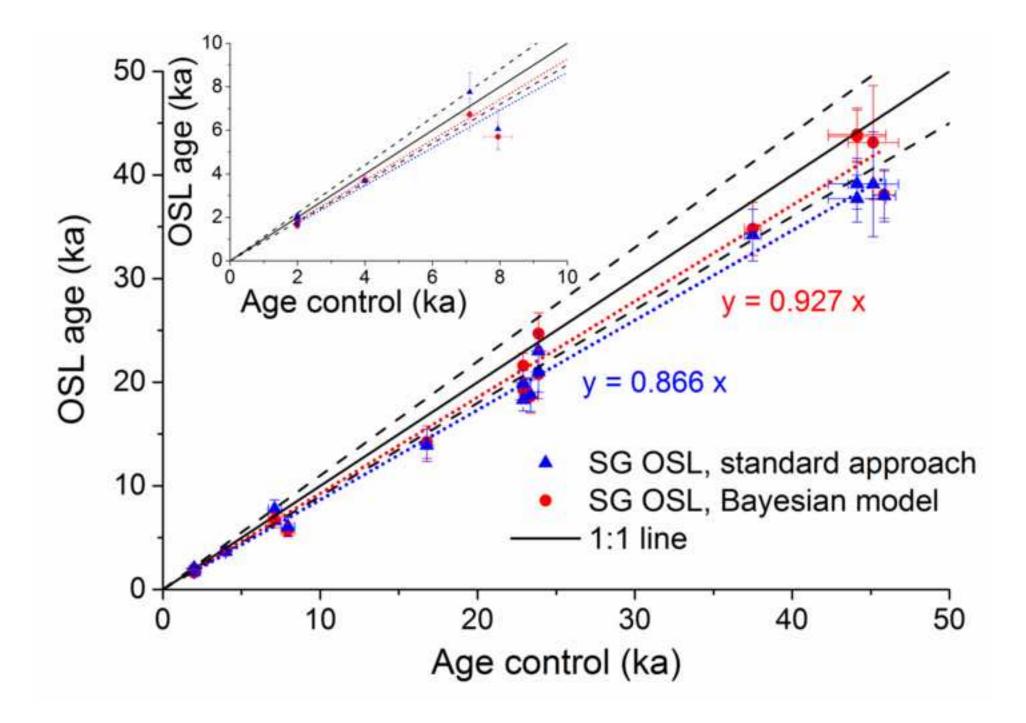
Fig. 2. Histogram of CAM overdispersion values determined for the samples studied here, and presumed to be well-bleached (see also Table 1).
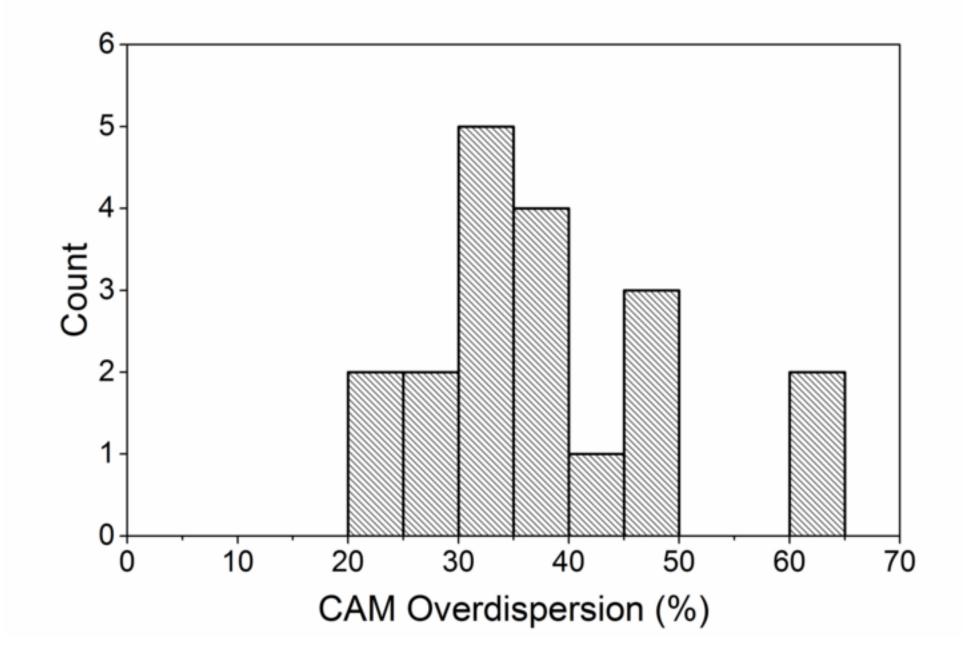
Fig. 3. Relative difference between single-grain OSL ages calculated with the standard approach, and reference ages, as a function of the same quantity when OSL ages are estimated with the Bayesian central dose model of Combès *et al.* (in press). The black line is the 1:1 line: points lying above the line indicate a better performance of the Bayesian model (Basar, for Bayesian SAR analysis: n=12); points below the line indicate a better performance of the standard approach (Analyst and CAM; n=7).
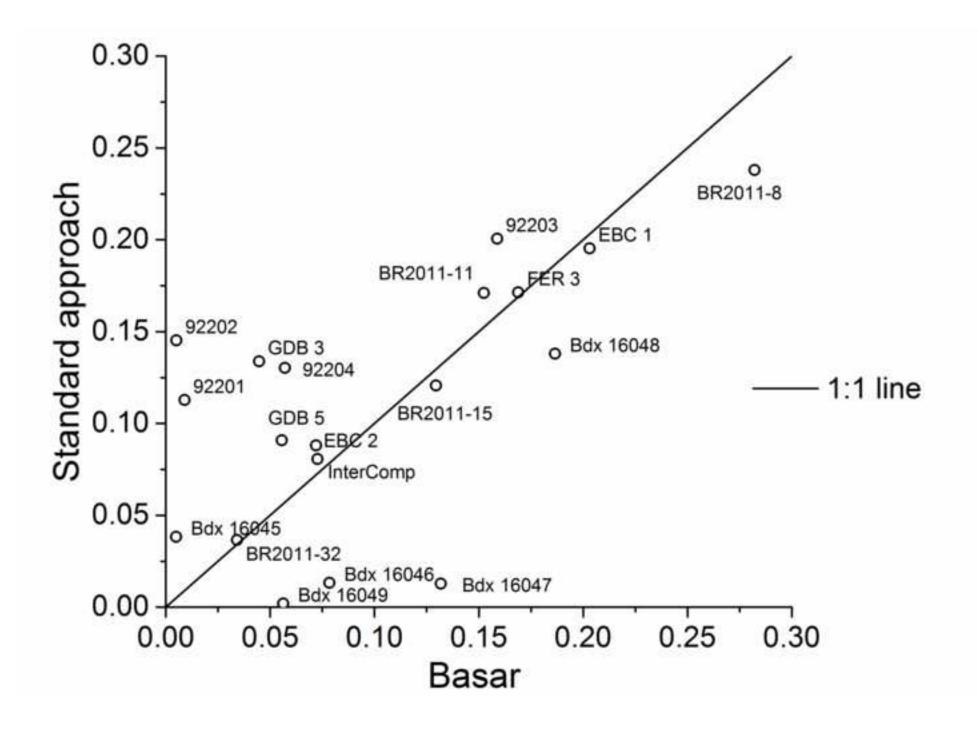
Fig. 4. Dose dependency of the discrepancies between the Bayesian central dose model of Combès *et al.* (in press) and the standard approach. (a) Relative difference between the ages obtained with the two analysis models as a function of dose; the dashed line corresponds to a linear fit to the data, and shows a positive correlation between this relative difference and dose. (b) Dose estimated with the Bayesian model as a function of CAM dose. The 1:1 line is represented by the black solid line.
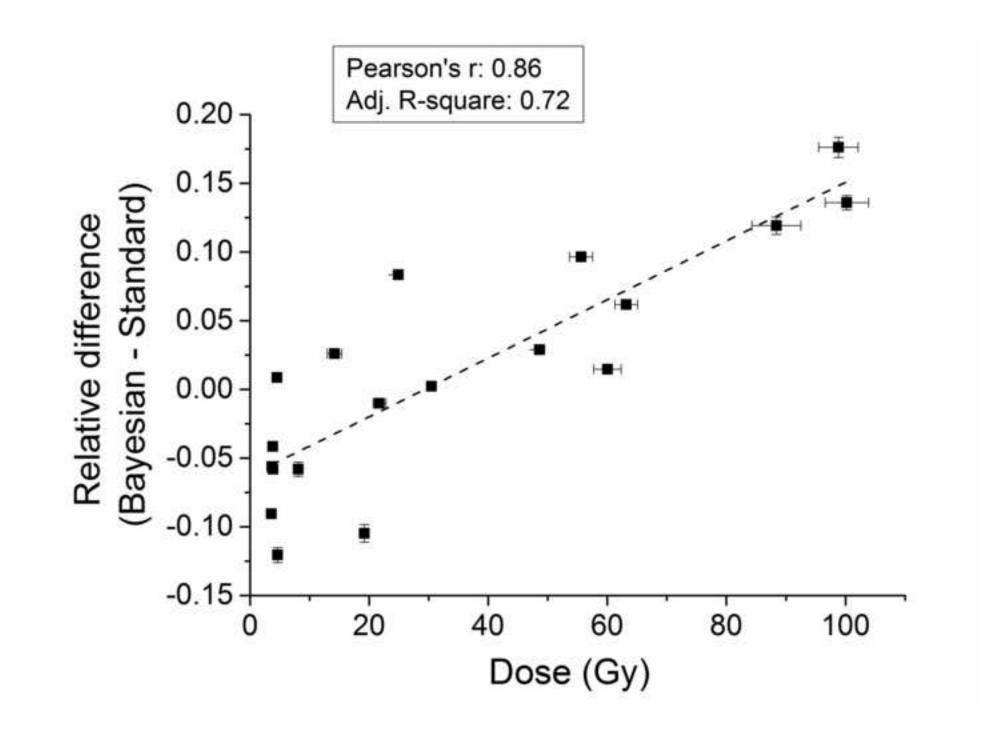
Fig. 5. Single-grain OSL to reference age ratio as a function of reference age: when the single-grain age has been estimated with the Bayesian central dose model of Combès et al. (in press; a), and with the standard approach (b).
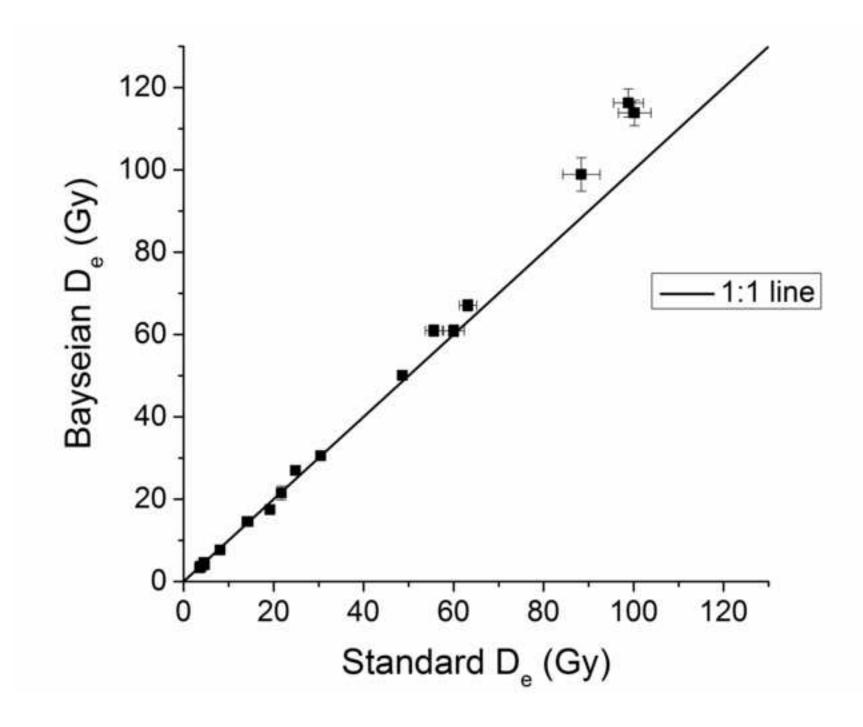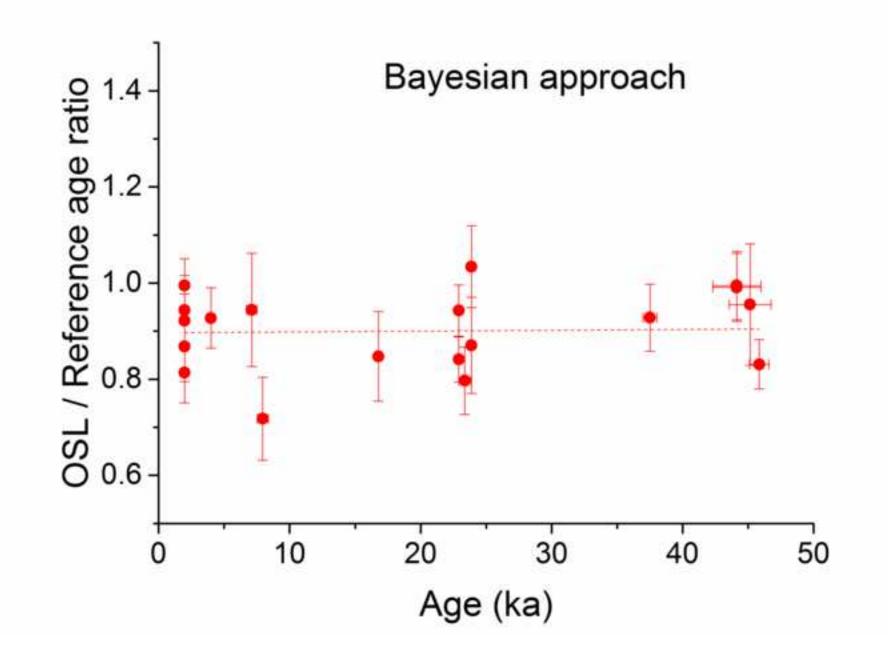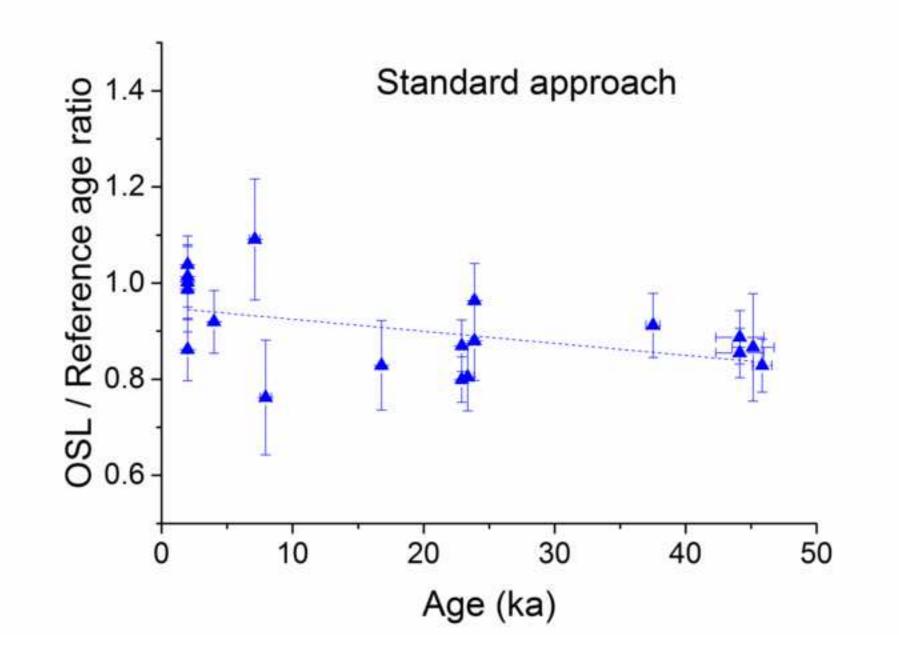
Figure 1b

Figure 2
Click here to download high resolution image

Bayesian approach

Standard approach

Table 1

Sample	Ref. age	Pseudo-sigma	Dose rate	σ	MG OSL Age (ka)	σ	Age ratio	σ	n	SG OSL, Bayesian Age (ka)	σ	Age ratio	σ	SG standard Age (ka)	σ	OD (%)	σ	Age ratio	σ
092201	44.1	1.8	2.56	0.12	42.5	2.2	0.96	0.05	273	43.7	2.5	0.99	0.07	39.1	2.5	48	3	0.89	0.07
092202	44.1	1.8	2.62	0.12	40.6	2.2	0.92	0.05	218	43.9	2.6	0.99	0.07	37.7	2.3	35	3	0.85	0.06
092203	22.9	0.2	3.45	0.16	21	1.2	0.92	0.05	218	19.3	1.1	0.84	0.05	18.3	1.1	36	2	0.80	0.05
092204	22.9	0.2	2.79	0.13	24.8	1.4	1.08	0.06	146	21.6	1.2	0.94	0.05	19.9	1.2	34	3	0.87	0.05
FER 3	45.8	0.7	1.58	0.08	42.2	2.8	0.92	0.06	190	38.1	2.3	0.83	0.05	38.0	2.5	45	3	0.83	0.06
InterComp	3.99	0.14	1.24	0.06	3.99	0.14	1.00	0.04	123	3.70	0.21	0.93	0.06	3.67	0.23	30	3	0.92	0.07
GDB 5	7.11	0.41	2.51	0.24					189	6.71	0.74	0.94	0.12	7.75	0.89	49	3	1.09	0.14
GDB 3	45.1	1.6	2.26	0.25	37.4	5.3	0.83	0.12	101	43.1	5.5	0.96	0.13	39.1	5.1	39	4	0.87	0.12
EBC 1	23.4	0.5	1.62	0.11	19.1	1.7	0.82	0.07	129	18.6	1.6	0.80	0.07	18.8	1.6	24	2	0.80	0.07
EBC 2	37.5	0.5	1.42	0.07	36.6	2.8	0.98	0.08	198	34.8	2.6	0.93	0.07	34.2	2.5	21	1	0.91	0.07
Bdx 16045	1.99	0.04	1.86	0.07					196	1.98	0.10	1.00	0.06	2.06	0.11	31	2	1.04	0.06
Bdx 16046	1.99	0.04	1.78	0.07					141	1.83	0.11	0.92	0.06	2.01	0.12	30	2	1.01	0.06
Bdx 16047	1.99	0.04	2.36	0.17					139	1.72	0.14	0.868	0.07	1.96	0.17	44	3	0.99	0.09
Bdx 16048	1.99	0.04	2.16	0.13					165	1.61	0.12	0.814	0.06	1.71	0.12	29	2	0.86	0.06
Bdx 16049	1.99	0.04	1.94	0.12					119	1.87	0.14	0.94	0.07	1.99	0.15	36	2	1.00	0.08
BR-2011-32	23.9	0.2	1.08	0.08	24.8	2.7	1.04	0.11	125	24.7	2.0	1.03	0.09	23.0	1.9	30	3	0.96	0.08
BR-2011-8	7.94	0.43	1.34	0.12	7.91	0.80	1.00	0.10	99	5.70	0.61	0.72		6.05	0.89	60	5	0.76	0.12
BR2011-15	23.9	0.2	1.04	0.08	23.8	2.62	1.00	0.11	68	20.8	2.4	0.87	0.10	21.0	1.95	33	4	0.88	0.08
BR-2008-11	16.8	0.1	1.02	0.07	19.3	2.0	1.15	0.12	62	14.2	1.6	0.85	0.09	13.9	1.6	62	6	0.83	0.09
Median							0.97					0.93						0.88	

Table 1. List of samples and ages analysed in the present study. 'Ref. age' and 'pseudo-sigma' stand for the age obtained independently from OSL measurements (see also Table S2, where details regarding this reference ages are provided). 'MG OSL' stands for multi-grain OSL; the ages have been calculated using either a plain average or the CAM, depending on the user. 'Age ratio' corresponds to the ratio of OSL to reference ages. 'n' is the number of grains analysed for each sample (after grain selection, as detailed in the original publications). 'SG OSL, Bayesian age model' corresponds to ages calculated with the Bayesian age model of Combès *et al.* (in press), while 'SG, standard age model' data results from the application of the CAM (Galbraith *et al.*, 1999) to data obtained with Analyst. 'OD (%)' corresponds to the overdispersion determined with the CAM. For more details on the list of samples, the reader is

referred to Aubry *et al.* (2012, 2014: samples 092201, -02, -03 and -04), Guérin *et al.* (2015a: sample Intercomp; 2015b: samples FER 3), Tribolo et al. (in prep., a: samples GDB 5, GDB 3; in prep., b: samples EBC1 and EBC2); Urbanova (in prep.: samples Bdx 16045 to 16049), Lahaye et al. (in press: samples BR-2011-32, -8, -15, -11).

e-component