1 **Towards an improvement of OSL age uncertainties: modelling OSL ages with systematic errors,**
2 **stratigraphic constraints and radiocarbon ages using the R package 'BayLum'**

3 Guillaume Guérin[1,2*], Christelle Lahaye[1], Maryam Heydari[1], Martin Autzen[3,4], Jan-Pieter Buylaert[3,4],
4 Pierre Guibert[1], Mayank Jain[3], Sebastian Kreutzer[5,1], Brice Lebrun[1], Andrew S. Murray[4], Kristina J.
5 Thomsen[3], Petra Urbanova[1], Anne Philippe[6].

6 *[1] UMR 5060 CNRS - Université Bordeaux Montaigne, IRAMAT-CRP2A, Maison de l'archéologie,*
7 *Esplanade des Antilles, 33607 Pessac cedex, France.*

8 *[2] Univ Rennes, CNRS, Géosciences Rennes, UMR 6118, 35000 Rennes, France*

9 *[3] Center for Nuclear Technologies, Technical University of Denmark, DTU Risø Campus, DK-4000*
10 *Roskilde, Denmark.*

11 *[4] Nordic Laboratory for Luminescence Dating, Department of Geoscience, Aarhus University, DTU Risø*
12 *Campus, DK-4000 Roskilde, Denmark.*

13 *[5] Geography & Earth Sciences, Aberystwyth University, Aberystwyth, Wales, United Kingdom*

14 *[6] Jean Leray Laboratory of Mathematics (LMJL), UMR6629 CNRS - Université de Nantes, France.*

17 **Keyword:**

18 OSL dating; Bayesian modelling; R package; Systematic errors; Covariance matrix; Stratigraphic
19 constraints

20 **Abstract**

21 Statistical analysis has become increasingly important in Optically Stimulated Luminescence (OSL)
22 dating since it has become possible to measure signals at the single grain scale. The accuracy of large
23 chronological datasets can benefit from the inclusion, in chronological modelling, of stratigraphic
24 constraints and shared systematic errors. Recently, a number of Bayesian models have been developed
25 for OSL age calculation; the R package 'BayLum' presented herein allows implementing different such
26 models, particularly for samples in stratigraphic order which share systematic errors. We first show
27 how to introduce stratigraphic constraints in 'BayLum'; then, we focus on the construction, based on
28 measurement uncertainties, of dose covariance matrices to account for systematic errors specific to
29 OSL dating. The nature (systematic versus random) of errors affecting OSL ages is discussed, based –
30 as an example – on the dose rate determination procedure at the IRAMAT-CRP2A laboratory
31 (Bordeaux). The effects of the stratigraphic constraints and dose covariance matrices are illustrated on
32 example datasets. In particular, the interest of combining the modelling of systematic errors with
33 independent ages, unaffected by these errors, is demonstrated. Finally, we discuss other common
34 ways of estimating dose rates and how they may be taken into account in the covariance matrix by
35 other potential users and laboratories. Test datasets are provided as supplementary material to the
36 reader, together with an R Markdown tutorial allowing ~~the user to reproduce~~the reproduction of all
37 calculations and figures presented in this study.

38

1. Introduction

Optically stimulated luminescence (OSL, called optical dating in Huntley *et al*., 1985) allows the dating of the last exposure of quartz grains to sunlight. The Single Aliquot Regenerative (SAR) dose protocol consists of comparing the natural luminescence signal to laboratory-generated signals induced by artificial irradiations (Murray and Wintle, 2000; Wintle and Murray, 2006). The corresponding measurements, in particular at the single-grain scale, result in large datasets characterised by significant scatter, owing to a number of dispersion factors (see, *e.g.* Thomsen *et al.*, 2005). An OSL age is then obtained by dividing the equivalent dose (i.e. in the case of coarse quartz grains, the dose absorbed by the mineral) by the dose rate to which quartz grains were exposed since the last exposure to light.

Statistical analysis, in geochronology, generally aims to improve the precision, accuracy, and/or range of dating methods. In the case of OSL dating, calibration errors on the laboratory source dose rate for natural dose estimation, and geochemical standards for dose rate assessment, have so far resulted in age uncertainties (at 1 sigma or 68 % confidence) of, at best, ~5% (see, *e.g.*, Duller, 2008; Guérin *et al*., 2013).

Note that in what follows, the unit of analysis is a sediment sample; we assume that each sample corresponds to a deposition event, and thus to a single age (no post-depositional mixing is considered). The system of analysis is the laboratory in which the measurements are performed and includes both the apparatus and associated calibration standards. It should be emphasised here that field equipment is part of what we call the laboratory; this is important for the definition of what we call systematic errors. By definition, an error is the difference between the measured or observed value of a physical quantity, and its true (but unknown) value. Thus, by systematic errors, we refer to random errors affecting equipment calibration: whereas each of these errors may be assigned a Gaussian probability density function with zero mean and a known variance (the square root of the variance being generally referred to as uncertainty), at the scale of the laboratory this error takes a fixed, unknown value that affects all measurements in the same direction. Of course, other sources of errors may exist (for example when using the infinite matrix assumption to calculate grain size attenuation factors, see *e.g.,* Guérin et al., 2012), but in this article, we consider only known, quantified sources of errors.

Over the past few years, several models for routine Bayesian analysis of SAR OSL and dose rate data were developed to reflect better, and take advantage of, the measurement procedures implemented to calculate OSL ages. Among those models, Combès *et al.* (2015) proposed one for calculating the central dose values for well-bleached samples, leading to higher overall accuracy (see Guérin *et al.*, 2015a) compared to the most commonly used model for OSL data analysis (the Central Dose Model: CDM, Galbraith *et al.*, 1999; note: we changed the original terminology following Galbraith and Roberts, 2012). Combès and Philippe (2017) developed models capable of dealing with individual and systematic multiplicative errors for OSL age calculation including stratigraphic constraints (for general introductions on a statistical analysis of OSL data, but also the statistical models discussed hereafter and associated prior distributions, the reader is referred to Combès *et al*., 2015; Combès and Philippe, 2017, and references therein).

To implement the Bayesian models of Combès *et al.* (2015) and Combès and Philippe (2017) in practice, and provide easy access to the community, an R package (R Core Team, 2020) named 'BayLum' (Christophe *et al*., 2020; version 0.2.0) has been developed and released on the Comprehensive R Archive Network (CRAN; see also Mercier *et al*., 2017, for a first implementation of the central dose model developed by Combès *et al*., 2015). First features of this 'BayLum' package

84   were presented by Philippe *et al.* (2019) and its performances, when one is confronted either with
85   large dose values or with dose variability issues, were tested in laboratory-controlled experiments
86   (Heydari and Guérin, 2018) and later applied to various case studies (Lahaye *et al*., 2018; Carter *et al*.,
87   2019; Heydari *et al*., 2020, 2021; Chevrier *et al*., 2020).

88   The purpose of this paper is to focus on the treatment of stratigraphic constraints and
89   systematic errors for chronological modelling using 'BayLum', *i.e.* it goes beyond than what was first
90   demonstrated by Philippe *et al.* (2019); together with the association of independent, more precise
91   ages (^{14}C in this work), such modelling is expected to reduce OSL age uncertainties. In the past, other
92   approaches to model systematic and random, individual errors in the field of palaeodosimetric dating
93   methods were proposed; in particular, Millard (2006a, 2006b) developed a Bayesian approach quite
94   close to that presented here, but which – among different other things (see Combès and Philippe,
95   2017, for a more detailed discussion) – is limited in its applicability.

96   Herein we present a Bayesian modelling case study. (1) We start with how data should be pre-
97   treated prior to using the 'BayLum' package; a simple example of chronological modelling (samples
98   considered independent, *i.e.* without stratigraphic constraints and shared errors) is first presented,
99   yielding an output from the 'BayLum' package to serve as a reference for the following, more elaborate
100   models. (2) In the next step, we detail how the user can integrate stratigraphic constraints and the
101   effect on the chronological inference. It should be noted that we take here the stratigraphic
102   information for granted, but we warn the user against treating such information lightly, as it bears
103   great consequences on the age calculation (cf. discussions in Heydari et al., 2020, 2021). (3) Then, most
104   importantly we explain how to build a dose covariance matrix in practice to take into account
105   systematic errors (for the definition of this matrix, the reader is referred to Combès and Philippe, 2017)
106   and what effect it has on a series of ages. (4) For this purpose, we base our approach on dose rate
107   measurements as performed by Guérin *et al.* (2015b) at the IRAMAT-CRP2A laboratory. The effect of
108   integrating independent data such as radiocarbon ages, which usually do not share systematic errors
109   affecting OSL data, is then illustrated. (5) Finally, we discuss different ways to measure dose rates and
110   various assumptions that can be made regarding the nature (systematic or random) of additional
111   sources of errors in OSL dating (see also Rhodes et al., 2003 for a similar discussion).

112   To help the reader, we provide as supplementary information an R markdown document with
113   commented lines of code and example datasets, so that everything presented here may be
114   reproduced.

2. Samples and methods

2.1. Case study

117   To illustrate how to model OSL ages, both in stratigraphic constraints and sharing systematic
118   errors, using the R 'BayLum' package, we use the data from two sediment samples (FER 1 and FER 3)
119   already dated by quartz OSL (Guérin *et al.,* 2015b). These samples were taken from the archaeological
120   site of La Ferrassie (France) and prepared following standard chemical preparation procedures applied
121   to luminescence-dating samples. While modelling with 'BayLum' may be applied to both multi-grain
122   and single-grain OSL datasets, in the following we only focus on single-grain data, as this is probably
123   where the need for appropriate statistical models is most acute (the reliability of multi-grain OSL has
124   been demonstrated when using a plain average (mean) for palaeodose estimation; see, *e.g.,* Murray
125   and Olley, 2002; for theoretical justification, see Guérin *et al.,* 2017). Single-grain OSL data were
126   measured using an automated Risø TL/OSL reader (DA 20) fitted with a single grain attachment (Duller
127   *et al.*, 1999; Bøtter-Jensen *et al.*, 2000). A standard SAR protocol (Murray and Wintle, 2000; 2003) was
128   used to measure single-grain equivalent doses, after checking its suitability for the samples under

129 investigation. A comparison between quartz OSL and feldspar IRSL signals for these two samples, as
130 well as comparison with radiocarbon, showed that these samples were well-bleached at the time of
131 deposition and unaffected by post-depositional mixing. As a result, the use of central dose models is
132 fully justified (it should be noted here that at the time of writing, 'BayLum' does not yet include the
133 Bayesian model of Christophe *et al.,* 2018, allowing the analysis of poorly bleached samples).

2.2. Data pre-treatment

135 The Bayesian modelling implemented in 'BayLum' requires information of different natures: (i)
136 raw OSL data in the form of BIN/BINX file(s), (ii) list(s) of grains to be included in the modelling (based
137 on pre-defined selection criteria, *e.g.* on recycling and/or recuperation ratios), (iii) file(s) indicating how
138 the data should be processed (signal integration channels, reproducibility of the instrument(s), etc.)
139 and (iv) both natural (in $Gy.ka^{-1}$) and laboratory (in $Gy.s^{-1}$) dose rates. Based on these data, the
140 calculations are performed all at once using Markov Chain Monte Carlo (MCMC) computations; as a
141 result, unlike in standard frequentist data processing, there is no succession of steps in data analysis
142 (for example, individual equivalent dose estimates are not parameterised, unlike when the CDM is
143 used). While Combès *et al.* (2015) argue that this results in a better statistical inference about the age
144 (or palaeodose), it also comes with a downside: the user cannot visualise the data during the statistical
145 analysis. In particular, the fact that the user must specify the list of grains to be included in the analysis
146 implies that one should always pre-treat the samples in a standard way, by using, *e.g. Analyst* (Duller,
147 2015) or the R 'Luminescence' package (Kreutzer *et al.*, 2012; Kreutzer *et al.*, 2020) to visually check
148 the data but also investigate the effect of various selection criteria on the datasets (see for example
149 Thomsen *et al.*, 2016, on the effect of applying various selection criteria when with frequentist
150 statistical models; see Heydari and Guérin, 2018, for a similar study in a Bayesian framework).

151 In other words, using 'BayLum' for age calculation should not, and does not, prevent the user
152 from a careful and critical examination of the measured OSL data. In particular, before running age
153 calculations using the 'BayLum' package, it is important that the user already has identified potential
154 problems – *e.g.*, saturation and/or dose rate variability (see Heydari and Guérin, 2018, for adapted
155 modelling solutions).

3. First simple model and output

157 We first ran the function `Generate_DataFile()` for the OSL samples FER 1 and FER 3,
158 with the same lists of grains as those used for age calculation by Guérin *et al* (2015b): all grains with
159 an uncertainty smaller than 20% on the first test dose signal were selected. A large number of grains
160 appeared to be in saturation for these samples (in *Analyst*, there is no intersection of the natural L/T
161 signal, or the sum of this sensitivity corrected natural signal and its uncertainty, with the dose-response
162 curve). As a result, following Thomsen *et al.* (2016), an additional selection criterion was added, based
163 on the curvature parameter of the dose-response curves. All grains for which the D_0 value, obtained
164 with Analyst as described by Guérin *et al*. (2015b), was smaller than 100 Gy, were rejected from the
165 analysis (note however that such a selection criterion may not be necessary when working with
166 'BayLum': Heydari and Guérin, 2018).

167 In practice, the data is contained in two folders named after the samples and provided as
168 Supplementary Material. Each folder contains one BIN/BINX-file (*i.e.* OSL measurements; note that
169 only a small fraction of the measured grains is included in the Supplementary Material) and four CSV-
170 files:

171 - 'DiscPos.csv' lists all selected grains;

172   - 'Rule.csv' gives the rules for generating L_x/T_x data (integration channels for both the natural
173   or regenerated and test dose signals, uncertainty arising from the reproducibility of the OSL
174   measurements, and number of SAR cycles to remove for curve fitting, if any - it may, for example, be
175   desirable to remove recycled points and/or IR depletion points);

176   -   'DoseSource.csv' gives the laboratory source dose rate and its variance;
177
178   -   'DoseEnv.csv' gives the dose rate to which the sample was exposed during burial.

179   We ran the function `AgeS_Computation()` with a prior age interval limited to between
180   10 ka and 100 ka for each sample (so that the bounds are far from the age values obtained using the
181   arithmetic mean of equivalent doses, namely 37 ± 2 ka and 40 ± 2 ka, respectively). The dose-response
182   curves were fitted, as in *Analyst* in our previous study, with single saturating exponential functions
183   passing through the origin. All uncertainties, affecting both environmental and laboratory dose rates,
184   were included in the calculation, as is common practice in luminescence dating; however, the
185   covariance of ages was not modelled here, so the results are equivalent to those one would obtain by
186   running subsequent individual age calculations for each of the two samples.

187   To run the `AgeS_Computation()` function, the user must choose a model for the
188   distribution of individual equivalent doses around the central dose; the different options are Cauchy,
189   Gaussian or lognormal distribution (in the latter case, the central dose may be estimated either by the
190   mean or the median of the distribution). A Cauchy distribution (sometimes also called Lorentz
191   distribution) is a symmetric distribution which was chosen by Combès et al. (2015) because it has heavy
192   tails, *i.e.* extreme values have a non-zero probability. Hence, the Cauchy distribution seemed to be
193   well-suited for the analysis of widely-dispersed datasets including outlier values such as single grain D_e
194   distributions.

195   Coming back to the samples from La Ferrassie, on top of saturation problems Guérin *et al*.
196   (2015b) also identified dose rate variability as an important factor of dispersion in equivalent doses:
197   the values of the CDM overdispersion parameter for the D_e distributions of the samples were equal to
198   29 ± 3 % and 35 ± 3 %, respectively. If we assume that this overdispersion arises from dose rate
199   variability to single grains of quartz, Heydari and Guérin (2018) using laboratory-controlled
200   experiments showed that the Cauchy distribution and the CDM should both lead to ~5-10% age
201   underestimation, because both models are biased. Consequently, we did not use the Cauchy
202   distribution model. Instead, we modelled the equivalent dose distribution by a lognormal distribution
203   (one could also have chosen a Gaussian function) from which the mean (rather than the median) was
204   used to estimate the central dose. Indeed, Guérin *et al*. (2017) formally demonstrated that the median
205   of the lognormal distribution (as used in the CDM) is a biased estimator and leads to age
206   underestimates when dose rates are dispersed.

207   After 5,000 iterations of 3 independent Markov Chains, we observed good convergence, as
208   seen in the Markdown document provided as supplementary material (for a discussion of the
209   convergence of the Markov Chains, the reader is referred to Philippe *et al*., 2019). The upper limit of
210   the 95% Confidence Intervals for the Gelman and Rubin indexes of convergence (Gelman and Rubin,
211   1992) were all smaller than 1.05, also indicating satisfying convergence of the 3 independent Markov
212   Chains (here again, the reader is referred to Philippe *et al*., 2019, who suggested 1.05 as the maximum
213   acceptable value). The obtained 95% Credible Intervals (C.I.) for the ages of samples FER 1 and FER 3
214   are [34.1; 43.3] ka and [36.6; 47.8] ka, respectively (Fig. 1; Table 1) and are consistent with the ages
215   obtained by Guérin *et al.* (2015b) with a much simpler approach (unweighted arithmetic mean of
216   equivalent doses). It should be emphasised here that the two 95% C.I. for ages overlap. Fig. 2 shows a

217 bivariate scatter plot of a sample of observations from the joint posterior distribution of the two ages,
218 as generated by the Markov Chains; in such a plot, each point corresponds to one realisation of the
219 ages of the two samples investigated in the MCMC. Fig. 3 shows the corresponding probability
220 densities for the ages estimated jointly, based on kernel density estimates (KDE), and the marginal
221 probability densities. No correlation is observed on the joint probability density, which is symmetrical
222 and bell-shaped. One can already compare the results obtained with this Bayesian model (lognormal-
223 average) for sample FER 3 with the radiocarbon ages obtained independently for the same layer by
224 Guérin *et al.* (2015b). The 95% C.I. for the 3 ^{14}C ages are bound by the interval [44.4; 47.3] ka, which
225 means that the OSL and radiocarbon ages are in good agreement, which was not the case when
226 calculating the ages with the CDM (38 ± 2 ka; this OSL age corresponds to ~15 % underestimation and
227 is broadly consistent, within uncertainties, with theoretical predictions stated above). Thus, even
228 without further modelling, the 'BayLum' lognormal-average model seems to provide OSL ages in better
229 agreement with radiocarbon.

230 **4. Stratigraphic constraints**

231 Samples FER 1 and 3 belong to two different stratigraphic layers: sample FER 1 (Layer 7) lies
232 stratigraphically above sample FER 3 (Layer 5B), so we know that the age of sample FER 1 must be less
233 than that of sample FER 3 (*i.e.* sample FER 1 is younger than sample FER 3; for detailed stratigraphic
234 information on the site of La Ferrassie, which is of paramount importance in this section, the reader is
235 referred to Guérin et al., 2015b). To encode this information, the function `AgeS_Computation()`
236 takes as argument the object `StratiConstraints`, which is a matrix whose size depends on the
237 number of analysed samples. First, the data in the DATA object (which is the output of the function
238 `Generate_DataFile()`) must be ordered in stratigraphic order from top to bottom: thus, in our
239 case the list of names used by the function `Generate_DataFile()` is FER 1, FER 3 (rather than FER
240 3, FER 1). Then, the stratigraphic matrix contains numbers equal to 0 or 1, indicating the applied
241 bounds to the age of each sample. The matrix contains a number of rows equal to the number of
242 samples plus one and a number of columns equal to the number of samples. The first row contains the
243 value 1 in each column, which indicates that the younger age bound specified as prior information (10
244 ka in our example, *cf.* section 3 above) when running the function `AgeS_Computation()` applies
245 to all samples. Then, for all *j* in {*2, ..., Nb_Sample+1*} and all *i* in {*j, ..., Nb_Sample*},
246 `StratiConstraints[j,i]`=1 if the age of sample whose number ID is equal to *j-1* is less than
247 the age of sample whose number ID is equal to *i*. Otherwise, `StratiConstraints[j,i]`=0. In
248 practice, in our case `StratiConstraints [1,]` = (1,1), `StratiConstraints [2,]` = (0,1)
249 (which means that sample FER 1 is not younger than itself but is younger than sample FER 3) and
250 `StratiConstraints [3,]` = (0,0) (which means that sample FER-3 is neither younger than
251 sample FER-1 nor than itself). Note: in the markdown document provided as Supplementary Material,
252 the corresponding code lines are commented to make this description easier to follow.

253 Running the function `AgeS_Computation()` with this matrix of stratigraphic constraints
254 only marginally affects the ages, in this case, the 95% C.I. become [34.3; 42.9] ka and [38.1; 48.5] for
255 samples FER-1 and FER-3, respectively (Table 1). One can also look at the bivariate scatter plot of
256 observations from the joint posterior distribution (Fig. 4): one can see that this scatter plot is truncated
257 in the upper left-hand corner – illustrating the fact that the age of sample FER 1 can never be greater
258 than that of sample FER 3 (see Fig. 2 for comparison). By contrast, the KDE estimate (Fig. 5) also shows
259 a positive correlation but does not reveal the truncation (whereas the stratigraphic constraint imposes
260 a null probability for all pairs of ages above the 1:1 line).

261

5. Dealing with multiple sources of errors through a covariance matrix

5.1. General considerations

In the previous calculations, all the variance is treated as random, whereas common, systematic errors affect all ages in the same direction, although to varying degrees (so systematic errors are unlikely to result in stratigraphic inversions). One of the main advantages of applying the models implemented in the 'BayLum' package – contrary to other chronological modelling tools such as *OxCal* (Bronk Ramsey and Lee, 2013) or *Chronomodel* (Lanos and Philippe, 2018) – lies in the possibility to include the structure of uncertainties specific to OSL dating. For instance, a radiocarbon age is derived only from the ratio of ^{14}C to ^{12}C (on top of which comes the more complex problem of calibration), whereas an OSL age involves a large number of measurements, each with its uncertainty (Aitken, 1985; 1998). The OSL measurements required for the determination of the palaeodose are relatively standardised through the widespread use of the SAR protocol (Murray and Wintle, 2000; Wintle and Murray, 2006). Conversely, there are several approaches – each with its equipment and standards – to determine the various dose rate components. Given that these dose rates derive from different types of radiation (alpha, beta, gamma and cosmic radiation) and are of various origins (mainly from potassium and the uranium and thorium decay chains), there are many more contributions to the age uncertainty from the dose rate term than from the palaeodose term, even though the size of the uncertainty on dose rate is of the same order of magnitude as that on palaeodose – see for example Murray *et al.*, 2015). As a result, there are almost as many ways of estimating systematic and random uncertainties as there are (combinations of) ways to determine dose rates; in any case, the notion of systematic error is only valid in a given context, which must always be made explicit. Combès and Philippe (2017) detailed the mathematical formulation of the dose covariance matrix, which links the ages of several samples measured using the same equipment and standards through common (systematic) errors (see also Philippe *et al.*, 2019). Nevertheless, the equations provided in this article are somewhat difficult to translate in practice; here, we propose to outline how we implement a covariance matrix adapted to (one example of) the measurements leading to OSL age calculation at the IRAMAT-CRP2A laboratory (Bordeaux). We emphasise that what follows is not prescriptive; it should be viewed as an example of a model of uncertainties. For alternative ways of estimating systematic and random errors, for example, due to different dose rates measurements, the reader is referred to the discussion (section 7.1).

Here, we consider the case of a series of *n* sediment samples taken from one unique site and all measured using the same equipment and standards. Let us consider the following relationship between palaeodoses, dose rates and ages (Combès and Philippe, 2017):

$$(D_1, \ldots, D_n) \sim \mathcal{N}\left((A_1\dot{d}_1, \ldots, A_n\dot{d}_n), \Sigma\right) \quad \text{(Eq. 1)}$$

where D_i is a random variable modelling the unknown palaeodose of sample *i*, $\mathcal{N}$ is the symbol for a Gaussian distribution, A_i is the unknown age estimate of sample *i* (that we are trying to determine), $\dot{d}_i$ the total dose rate to which this sample was exposed since burial ($\dot{d}_i$ is the observed dose rate, *i.e.* the result of the measurements) and Σ is the dose covariance matrix (for the full definition of the model, we refer the reader to Combès and Philippe, 2017). This covariance matrix verifies, for all (i,j):

$$\Sigma_{i,j} = A_i A_j \theta_{i,j} \quad \text{(Eq. 2)}$$

where θ is the matrix that the user needs to specify to run the calculations with 'BayLum'. It should be noted here that by default when running age calculations with 'BayLum', the off-diagonal elements are set to zero, *i.e.* the covariance in ages is not modelled.

305 Before entering the details specific to luminescence dating, let us consider a simple example of two
306 measurements $y_1 = \mu_1 + e_1 + f$ and $y_2 = \mu_2 + e_2 + f$ where μ_1 and μ_2 are fixed measurands and e_1, e_2 and
307 f are all independent random errors from distributions with mean zero. The covariance of y_1 and y_2 is
308 the variance of f (so the off-diagonal elements of the matrix are equal to this variance). For each
309 sample, the diagonal element of the corresponding covariance matrix is the sum of all the components
310 of variance for that sample. The variety of physical quantities to measure to determine dose rate, and
311 their relationship with the dose rate contributions, will now be discussed with this simple definition in
312 mind.

5.2. Implementation in practice

314 First, we detail the series of measurements carried out, and we introduce the corresponding
315 notations for the estimates and associated uncertainties. Table 2 summarises all physical units and
316 associated error standard deviations; as a general rule, we assume that all error terms are Gaussian
317 variables with the expected value (mean) equal to zero and a fixed, known standard deviation (see for
318 example Eq. 2 in Combès and Philippe, 2017). For clarity, in the following relative standard deviations
319 are described by the letter σ, while absolute standard deviations are denoted by s; moreover, each
320 standard deviation corresponding to random errors (*i.e.*, when the error varies from sample to sample)
321 is identified by the letter i in the subscript. The absence of this letter in the subscript indicates that the
322 measurement error affects all samples.

5.2.1. Equivalent doses and OSL measurements

324 Equivalent doses are determined from OSL measurements performed on a luminescence
325 reader equipped with a radioactive beta source, whose dose rate and associated relative standard
326 deviation of the error, noted $\dot{d}_{\mathrm{lab}}$ and σ_{lab}, are known. There are several ways the latter term can be
327 determined; in its simplest form, it includes the standard deviation of the error on the absolute dose
328 absorbed by the standard reference material (in our case calibration quartz provided by DTU Nutech,
329 *cf.* Hansen *et al.,* 2015) and an error term due to replicate measurements of several aliquots of this
330 calibration material. Using a large number of measurements repeated in time, as suggested by Hansen
331 *et al.* (2015), may somewhat complicate the matter, but this goes beyond the scope of the present
332 study.

333 In practice, regeneration doses are delivered by irradiating the aliquots for a given duration (in
334 s). This duration is converted to absorbed energy dose (Gy) by multiplication with the source dose rate
335 (Gy.s^{-1}). Strictly speaking, the error on the source dose rate affects all regeneration doses, and so this
336 error term should appear in the dose/luminescence relationship (right side of the directed acyclic
337 graph shown in Fig. 7 of Combès and Philippe, 2017). However, it is common practice in luminescence
338 dating to first calculate an equivalent dose in seconds of irradiation for each aliquot, then convert this
339 to Gy and calculate an average (or determine another central parameter such as with the CDM), and
340 only then consider σ_{lab}. This is what led, *e.g.*, Jacobs et al. (2008), to exclude the associated standard
341 deviation from the total OSL age uncertainties, to test the assumption of a time gap between two
342 series of ages. Here, for simplicity, we take the same route, and hence the relative error on the
343 laboratory source dose rate becomes a relative, systematic error on the equivalent doses.

344 One may thus write that the error on the dose D_i arising from the calibration of the source
345 follows a Gaussian distribution with mean 0 and variance $(\sigma_{\mathrm{lab}}\, A_i \dot{d}_i)^2$.

346

5.2.2. Dose rates

348   When it comes to the dose rate term, here we restrict ourselves to the case of coarse quartz
349   grains measured after HF etching to remove the alpha dose rate component: the total natural dose
350   rate is the sum of an internal dose rate, external beta and gamma dose rates, and cosmic dose rates.

351   **Cosmic dose rates**

352   We consider that cosmic dose rates are determined following Prescott and Hutton (1994)
353   based on the burial depth of the dated samples, which may be different from the present-day thickness
354   of the overburden. As a result, the error on cosmic dose rate estimates depends on the error
355   estimation of this effective burial depth since the dated sediment was deposited. Because the
356   relationship between cosmic dose rates and burial depths is not linear, and because the error on this
357   burial depth may not be systematic (*e.g.* in cases where successive, yet of unknown duration, erosion
358   and deposition events happened between the deposition of superimposed sedimentary layers – see
359   Aitken, 1998, p. 65, for a discussion) even at the scale of a site the error associated to cosmic dose
360   rates cannot easily be treated as systematic. For *i={1,…,n}*, $\dot{d}_{\mathrm{cosmic},i}$ and $s_{\mathrm{cosmic},i}$ denote the estimate
361   of the average cosmic dose rate to which sample *i* has been exposed and its associated standard
362   deviation.

363   **Beta dose rates**

364   We consider the beta dose rates as determined from concentrations (or activities) of ^{40}K and
365   in radioelements from the U- and Th- decay chains, converted to dose rates using specific conversion
366   factors (*e.g.,* Guérin *et al.*, 2011). At the IRAMAT-CRP2A laboratory, these concentrations are usually
367   determined with low-background, high-resolution gamma-ray spectrometry following Guibert and
368   Schvoerer (1991). The simplest case is that of ^{40}K, since only one peak is used (at 1.461 MeV); the
369   concentration in sample *i*, denoted [K]$_i$ is equal to the concentration in the standard multiplied by the
370   ratio in count rates (the count rate observed for the investigated sample is divided by the count rate
371   observed for a reference material). Thus, we consider in this paper that the standard deviation of the
372   error on the ^{40}K concentration includes three components: the standard deviation of the error on the
373   concentration in the standard, and the counting uncertainties both on the standard and on the
374   measured sample. The counting uncertainties are calculated, assuming Poisson statistics. Of these
375   three sources of errors, only one is treated as random – namely the counting uncertainty of the sample;
376   the other two standard deviations (corresponding to the counting of the standard and to the error of
377   the radioelement concentration in the standard) are quadratically summed and considered as a
378   systematic source of error. One considers for sample *i* the beta dose rate from potassium $\dot{d}_{\beta,\mathrm{K},i}$ – after
379   correction for grain size-dependent attenuation using the factors from Guérin *et al.*, (2012a); and for
380   moisture content following Nathan and Mauz (2008) (see the discussion section below regarding
381   uncertainties on these correction factors). Neglecting uncertainties in the dose rate conversion factors,
382   we call $\sigma_{\mathrm{K},i}$ the relative random standard deviation of the error on the ^{40}K concentration; its systematic
383   counterpart σ_{K} is common to all samples. It should be emphasised here that systematic errors on
384   radioelement concentrations, although being shared by all samples, will affect all ages in the same
385   direction but not necessarily by the same amount (even in relative terms, contrary to the error on
386   laboratory beta source calibration) because the relative contribution of beta dose rate from potassium
387   to the total dose rate may vary from one sample to another. The beta dose rates from the U- and Th-
388   series come from a number of radioelements in the corresponding chains; here, for simplicity we
389   consider each series to be in secular equilibrium (this is generally the case for ^{232}Th but may not be
390   true for the U-series, see, *e.g.* Guibert *et al.*, 1994; 2009; Lahaye *et al.,* 2012). Thus, for each sample,
391   the concentrations in ^{238}U and ^{232}Th are converted to dose rate contributions denoted $\dot{d}_{\beta,\mathrm{U},i}$ and
392     $\dot{d}_{\beta,\mathrm{Th},i}$. In contrast to the case of ^{40}K, the analysis of the high-resolution spectra for these radioactive

393 chains is based on a number of primary gamma rays (whereas there is only one ray for K); more
394 specifically, a weighted mean of the concentrations determined from each ray included in the analysis
395 (after taking interference into account) is calculated to estimate the concentration of U
396 (respectivelyand that of Th). As a result, the standard deviation of the error on the concentration in U
397 (resp.and that of Th) in the sample comes from two sources: the relative standard deviation on the
398 concentration of the standard corresponds to a systematic error and is denoted σ_U for U (resp. σ_{Th} for
399 Th); conversely, the other relative standard deviations (arising from the counting of the standards and
400 of the sample) are treated as random and quadratically summed to obtain $\sigma_{U,i}$ (resp.and $\sigma_{Th,i}$).

Internal dose rates

402 Unless the internal radioelement concentration is experimentally determined (in which case
403 one needs to consider both systematic and random sources of error for each sample, as is done for
404 beta dose rates), some have suggested using a fixed internal dose rate of 0.06 ± 0.03 Gy.ka^{-1} (Mejdahl,
405 personal communication to Murray, based on Mejdahl, 1987). In this case, we may assume that the
406 dated quartz grains are all of the same origin, and have the same internal radioelement concentration;
407 as a result, we associate a systematic standard deviation s_{int} with the internal dose rate $\dot{d}_{int}$.

Gamma dose rates

409 Gamma dose rates $\dot{d}_{\gamma,i}$ may be determined, as beta dose rates, from K, U and Th
410 concentrations in the sediment. In this case, the reader is referred to the corresponding section above.
411 However, it is relatively frequent, in the case of heterogeneous configurations at the 10 cm scale, that
412 gamma dose rates received by the samples do not correspond to the infinite matrix gamma dose rates
413 of the samples (see for example large gamma dose rate variations at the interface between sediment
414 and bedrock in a cave reported by Guérin *et al.*, 2012b: Fig. 7). In such contexts, gamma dose rates
415 may be determined by *in situ* measurements with Al$_2$O$_3$:C artificial dosimeters: these dosimeters are
416 measured with green-light stimulation and their calibration is based on a block of homogeneous bricks
417 located in the basement of IRAMAT-CRP2A (Richter *et al.*, 2010; Kreutzer *et al.*, 2018; note: we also
418 discuss below – section 7.1 – the use of portable spectrometers). Two sources of relative errors are
419 taken into account: a random standard deviation ($\sigma_{\gamma,i}$) accounting for measurement uncertainties, and
420 a shared calibration error including both standard deviations on (i) the true gamma dose rate in the
421 block of bricks and on (ii) the measurement of the dosimeters irradiated inside the block for calibration
422 of the source (σ_γ).

Water content

424 To account for the effect of water on dose rates, one commonly considers the following
425 equation (Zimmerman, 1971; Aitken, 1985):

$$\dot{d}_{\beta,i} = \frac{\dot{d}_{\beta,i,\mathrm{dr}\gamma}}{1+x_\beta WF_i}, \qquad \text{Eq. (3)}$$

427 where $\dot{d}_{\beta,i,\mathrm{dr}\gamma}$ is the beta dose rate in the dry sediment, WF_i represents the effective mass fraction of
428 water in the sediment during burial, and x_β is a water correction coefficient accounting for the fact
429 that water absorbs more beta dose than typical sedimentary elements, due to lower atomic numbers
430 (Nathan and Mauz, 2008). A similar equation applies to gamma dose rates, with a corresponding factor
431 x_γ (see Guérin and Mercier, 2012). The determination of the water content in the sediment over time
432 is a challenging task as it involves many different parameters, including past rainfall – see for example
433 Nelson and Rittenour (2015) for a discussion on how to determine water contents depending on
434 sediment grain size, hydrometric regimes, etc. One commonly employed solution is to measure the

water content at the time of sampling and assume it to be representative of that in the past; measuring the water content at saturation may then be a solution to evaluate an upper limit to this value; and depending on the context one may also propose a lower limit to the water content. One then obtains a way of quantifying the standard deviation of the error on the water content, although necessarily imperfect. Neglecting uncertainties on the water correction factors (x_β and x_γ) and calling $s_{WF,i}$ the absolute standard deviation of the mass fraction WF_i for sample i, one may write:

$$s_{\beta,H_2O,i} = \dot{d}_{\beta,i}\frac{s_{WF,i}}{1+x_\beta WF_i} \qquad \text{(Eq. 4)}$$

where $s_{\beta,H_2O,i}$ is the standard deviation of the beta dose rate for sample i due to the uncertainty on its water mass fraction.

Similarly, one may write:

$$s_{\gamma,H_2O,i} = \dot{d}_{\gamma,i}\frac{s_{WF,i}}{1+x_\gamma WF_i} \qquad \text{(Eq. 5)}$$

where $s_{\gamma,H_2O,i}$ is the standard deviation of the gamma dose rate for sample i due to the uncertainty on its water mass fraction. As a result,

$$s_{\gamma,H_2O,i} = \frac{\dot{d}_{\gamma,i}}{\dot{d}_{\beta,i}}\frac{1+x_\beta WF_i}{1+x_\gamma WF_i}s_{\beta,H_2O,i}. \qquad \text{(Eq. 6)}$$

To simplify the following equations, which are meant to be those used in practice, we introduce the relative standard deviation of the beta dose rate due to water content errors ($\sigma_{\beta,H_2O,i}$) and a parameter called λ_i defined by:

$$\lambda_i = \frac{1+x_\beta WF_i}{1+x_\gamma WF_i} \qquad \text{(Eq. 7)}$$

Finally, it should be emphasized that uncertainty on water content may well correspond to errors which are neither really random nor really systematic; in our view different modelling choices may be put forward and implemented, depending on the particular sedimentological and pedological context.

The θ matrix

With these considerations in mind on errors and their nature, the corresponding θ matrix (Eq. 21) to model these uncertainties is a square matrix containing one line (and column) per sample. The diagonal elements correspond to the sum of a term arising from the error on the laboratory source dose rate ($\dot{d}_i^2\sigma_{lab}^2$) and the total dose rate variance for each sample, for each i:

$$\theta_{i,i} = \dot{d}_i^2\sigma_{lab}^2 + s_{cosmic,i}^2 + \dot{d}_{\beta,U,i}^2(\sigma_{U,i}^2 + \sigma_U^2) + \dot{d}_{\beta,K,i}^2(\sigma_{K,i}^2 + \sigma_K^2) + \dot{d}_{\beta,Th,i}^2(\sigma_{Th,i}^2 + \sigma_{Th}^2) + s_{int}^2 +$$

$$\dot{d}_{\gamma,i}^2(\sigma_{\gamma,i}^2 + \sigma_\gamma^2) + (d_{\beta,U,i} + d_{\beta,K,i} + d_{\beta,Th,i} + \lambda_i d_{\gamma,i})^2\sigma_{\beta,H_2O,i}^2. \qquad \text{(Eq. 8)}$$

This long list of variance terms may seem rather complicated. However, it corresponds to the total variance arising from the laboratory beta source calibration, the errors on cosmic dose rates, environmental beta dose rates, internal dose rates, gamma dose rates, and finally the error arising from uncertainties in water content. In other words, we can also write

$$\theta_{i,i} = \dot{d}_i^2\sigma_{lab}^2 + s_{\dot{d}_i}^2 \qquad \text{(Eq. 92)},$$

where $s_{\dot{d}_i}^2$ is the variance of the dose rate to which sample i was exposed to during burial (it is the square of the uncertainty appearing next to the dose rate value in every luminescence dating article;

471 in our example, this term is the second one in the files DoseEnv.csv provided in Supplementary
472 Material).

473   Then, for $i \neq j$:

474 $$\theta_{i,j} = \dot{d}_{\gamma,i}\,\dot{d}_{\gamma,j}\sigma_\gamma^2 + \dot{d}_{\beta,U,i}\dot{d}_{\beta,U,j}\sigma_U^2 + \dot{d}_{\beta,K,i}\dot{d}_{\beta,K,j}\sigma_K^2 + \dot{d}_{\beta,Th,i}\dot{d}_{\beta,Th,j}\sigma_{Th}^2 + s_{int}^2 + \dot{d}_i\,\dot{d}_j\sigma_{lab}^2 \quad \text{(Eq. } 10\cancel{3}\text{),}$$

476 which characterises the amount of correlation between the doses of samples *i* and *j*, multiplied by their
477 ages. The θ matrix, like the dose covariance matrix Σ, is a symmetric matrix. The diagonal members
478 correspond to individual variances, while the non-diagonal terms express the fact that systematic,
479 shared errors link the measurements of the series of samples. As a result, running the functions
480 `AgeS_Computation()` and `Age_OSLC14()` with a θ matrix in which all non-diagonal members
481 are set to zero would be equivalent to running the same functions without the correlation matrix, or
482 running the function `Age_Computation()` independently for each sample – in which case all
483 sources of error are treated as random.

5.3. Examples

5.3.1. An illustrative, simplistic example without stratigraphic constraints

486   For illustration purposes, first, we did not apply stratigraphic constraints. We started with a
487 simplistic θ matrix containing in the diagonal the real error variances (Eq. 9$\cancel{2}$) as determined by Guérin
488 *et al.* (2015); the σ_{lab} value was equal to 0.02 (2% relative standard deviation of the calibration of the
489 laboratory beta source). The simplification comes from the off-diagonal members, for which in Eq.
490 (10$\cancel{3}$) we set all *s* and σ values equal to 0, except for the σ_{lab} value, set to 0.05. Obviously, this is not
491 self-consistent, but it corresponds to (i) random and systematic errors of approximately the same
492 magnitude (in practice, these two sources of errors are of the same order of magnitude – a few %) and
493 (ii) the simplest form of systematic errors. Indeed, in such a case, all ages are affected by the same
494 relative amount in the same direction.

495   Here again, after 5,000 iterations of 3 independent Markov Chains, we observed good
496 convergence. The obtained 95% C.I. are [33.9; 43.8] and [36.7; 48.1] ka for samples FER 1 and FER 3,
497 respectively. Fig. 6 shows bivariate scatter plots corresponding to the sampling of the Markov Chains
498 for the ages of samples FER 1 and FER 3 (which are calculated simultaneously) and Fig. 7 displays the
499 KDE together with the marginal probability densities. This set of figures illustrate the reason for the
500 generation of the two types of figures: the bivariate scatter plot is most appropriate for visualising the
501 effect of stratigraphic constraints (Fig. 4 above), whereas probability density figures best illustrate the
502 effect of modelling systematic errors. Indeed, as can be seen, there is a positive correlation between
503 the ages of samples FER 1 and FER 3: the greater the age of sample FER 1, the greater is the mean age
504 of sample FER 3. In other words, if the age of sample FER 1 were underestimated, then in all likelihood,
505 so would be the age of sample FER 3. Furthermore, the length of the C.I. for the age of each sample is
506 slightly larger than without modelling the covariance (*cf*. Table 1), *i.e.* modelling the covariances
507 slightly increases the age uncertainties. However, the positive correlation of ages has other, direct
508 consequences.

509   First, let us suppose that we have no knowledge of a stratigraphic link between the two
510 investigated samples, and wish to test the hypothesis that sample FER 1 is younger than sample FER 3.
511 The credibility of such an assumption can be tested using the function `MarginalProbability()`
512 of the 'Archaeophases' R package (Philippe and Vibet, 2020) devoted to the analysis of MCMC chains
513 for chronological inference. Without using the covariance matrix, the credibility of this hypothesis is

514 0.83; with the simplistic θ matrix, the credibility becomes 0.94; in other words, modelling the age
515 covariance reflects more faithfully the measurements and their uncertainties for such tests.

516 The second consequence concerns the duration of a hypothetical phase that would encompass
517 the deposition of sample FER 1 and that of sample FER 3. Indeed, since the ages vary together in the
518 MCMC, the duration of such a phase should be smaller when modelling the covariance than when all
519 the variance in ages is treated as random. Indeed, we could verify this assertion using the function
520 `PhaseStatistics()` of 'ArchaeoPhases' (Philippe and Vibet, 2020): with the simplistic covariance
521 matrix, the 95 % C.I. for the duration of this phase is [-1.4; 9.7] ka, whereas it is [-0.6; 7.6] ka when the
522 ages are calculated using the simplistic θ matrix.

523 **5.3.2. A real example, including stratigraphic constraints**

524 In a real case, since the relative contributions of the different dose rate components vary from
525 one sample to another, the correlation will be less pronounced. For more realistic calculations of the
526 ages of samples FER 1 and FER 3, we took the same values as above for the diagonal terms of the θ
527 matrix (Eq. 92); on the other hand, for the non-diagonal, covariance terms, we used the following
528 values: $\sigma_{lab} = 0.02$ (which corresponds to the experimentally determined calibration standard
529 deviation, including the uncertainty of the dose delivered to calibration quartz; Hansen *et al*., 2015),
530 $\sigma_K = 0.012, \sigma_U = 0.007$, $\sigma_{Th} = 0.007$ (for these values, which also include counting of the standards
531 used, the reader is referred to Guibert *et al.*, 2009; Guibert, 2002), and $s_{int} = 0.003$ Gy.ka^{-1}. We
532 provide as Supplementary Information a calculation spreadsheet allowing to build the covariance
533 matrix, intended for adaptation to the user-specific needs.

534 At the site of La Ferrassie, the uncertainties associated with the gamma dose rate observations
535 are more complex. Al$_2$O$_3$:C dosimeters were placed at the end of 25 cm long aluminium tubes and
536 inserted horizontally in the stratigraphic section at the location of sediment sampling. In an ideal case,
537 sediment should be uniform in a horizontal plane; however, for samples FER 1 and FER 3 only a rather
538 thin layer of sediment remained against the cliff wall (the layers of the sample were not present at the
539 site in any other location), which resulted in the dosimeters being inserted either in the karstic cliff
540 (the limestone contains little radioelements compared to the sediments, as shown in Fig. 5 of Guérin
541 *et al.*, 2015b) or at the interface between the cliff and the sediment. As a result, we took for $d_{\gamma,i}$ the
542 average between the gamma dose rates measured *in situ* (which underestimate the real gamma dose
543 rate because the effect of the cliff is over-represented) and the gamma dose rates derived from the K,
544 U and Th concentrations in the samples. The associated standard deviation, $\sigma_{\gamma,i}$, was calculated as the
545 difference between these two extreme values divided by 4, so that the 95% C.I. covers all possible
546 values. As this standard deviation is much larger than the analytical uncertainties, we neglected the
547 latter and considered $\sigma_{\gamma,i}$ to characterise random sources of errors since each sample has a different
548 environment and may be more or less far from the cliff.

549 The samples FER 1 and FER 3 are directly above and below, respectively, the Châtelperronian
550 layer at the site (layer 6). Sample FER 2 from this layer being poorly bleached, it is at present impossible
551 to model with 'BayLum'. However, an alternative to estimate the age of FER 2 consists of supposing
552 that it has a uniform prior probability density between the ages of samples FER 1 and FER 3:

553 $$P(A_2|data) \sim \iint \frac{\mathbb{I}_{[A_1;A_3]}}{A_3 - A_1} \pi(A_1, A_3|data) dA_1 dA_3 \qquad \text{(Eq. 11)}$$

554 where A_i is the age of sample *i*, $\mathbb{I}_{[A_1;A_3]}$ is the indicator function between A_1 and A_3, and
555 $\pi(A_1, A_3|data)$ is the posterior joint density of A_1 and A_3 knowing the data (*i.e.* the density estimated
556 with 'BayLum'). Doing so (see the markdown file for the corresponding code lines), working from the

557 output of 'BayLum' one obtains a 95% C.I. of [36; 46] ka, which can be compared with the confidence
558 interval of [36; 48] ka obtained by Guérin et al. (2015) with minimum age modelling.

559 **6. Integration of independent chronological data (radiocarbon)**

560 The 'BayLum' package also offers the possibility to include radiocarbon ages in the chronological
561 models (Philippe *et al.*, 2018); more specifically, radiocarbon ages are calibrated within 'BayLum', using
562 the function `AgeC14_Computation()` or `Age_OSLC14()` (in the latter case the function
563 necessitates at least one OSL age calculation). Introducing covariance matrices to account for
564 systematic errors on OSL data does not reduce the OSL age uncertainties; however, it becomes
565 particularly useful to correct for estimation biases when more precise ages, unaffected by these
566 systematic errors, are integrated into the models. To illustrate this, we decided to construct two
567 models constraining the age of FER 3; for illustration purposes, in this section, we used the simplistic
568 θ matrix described above in section 5.3.1. In the first case, we constrained the age of this sample by
569 imposing that a 'young' radiocarbon age (young compared to the age of sample FER 3 considered
570 alone) has an age greater than sample FER 3. In practice, we arbitrarily took a radiocarbon age of
571 38,000 ± 400 BP, which corresponds to [37.6; 39.9] ka cal. BP (95% C.I. using the IntCal20 curve, Reimer
572 *et al.*, 2020; the calibration was performed using 'BayLum', see Philippe et al., 2018). Naturally, the
573 credible intervals (both 68% and 95%) for sample FER 3 are shifted towards younger age values (*cf.*
574 truncation of the scatter plot illustrated in Fig. 3). So do the credible intervals for sample FER 1, since
575 the ages of the two OSL samples are close to each other even when considered independently of
576 radiocarbon data (in other words, the radiocarbon age 'pushes' the age of sample FER 3, which in turn
577 'pushes' the age of sample FER 1). In practice, the 95% C.I. become [33.3; 41.2] ka and [36.9; 42.3] ka
578 for samples FER 1 and FER 3, respectively. It can be noted here that in such a case the precision of the
579 age of sample FER 3 is increased (*i.e.* the length of the C.I. is much smaller than without the constraining
580 radiocarbon age). More interestingly, in the second case, we constrained the age of sample FER 3 by
581 imposing that an 'old' radiocarbon age (old compared to the age of sample FER 3 considered alone)
582 has an age younger than sample FER 3. In practice we – again, arbitrarily – took a radiocarbon age
583 equal to 44,000 ± 400 BP, which corresponds to [45.4; 47.4] ka cal. BP (95% C.I.). Here again, the effect
584 on the age of sample FER 3 is straightforward: the credible intervals are shifted towards older ages
585 (the 95% C.I. for the age of sample FER 3 becomes [45.7; 51.2] ka). Perhaps less intuitive is the effect
586 on the age of sample FER 1, which is not directly constrained by radiocarbon: because the ages of the
587 three samples are estimated jointly, and because of the systematic errors on the OSL ages, the age of
588 sample FER 1 is also shifted towards older ages: the corresponding 95% C.I. becomes [36.7; 45.8] ka.

589 **7. Discussion**

590 **7.1. Differing ways of estimating dose rates**

591 Every laboratory uses its specific equipment and calibration standards; if similar equipment as
592 described above is used, then only the values of the different terms need be changed. This case is
593 particularly relevant for equivalent dose measurements, and hence the term σ_{lab} associated with $\dot{d}_{lab}$.
594 Conversely, for dose rate determination, several other experimental devices and techniques are
595 commonly used. If beta and/or gamma dose rates are determined based on the determination of
596 concentration in K, U and Th, (for example by mass spectrometry, neutron activation, etc.), then the
597 situation is similar as that described for beta dose rates above.

598 Counting techniques (alpha, beta, and gamma in the case of the threshold technique: Løvborg
599 *et al.*, 1974) may also be used for beta and gamma dose rate estimation. In the case of beta counting,
600 the conversion factor from count rate to dose rate depends on the emitting radioelement

601 (Ankjærgaard and Murray, 2007; see also Cunningham *et al.*, 2018). This dependency is a source of
602 error that may not easily be characterised by a systematic error (so there is no contribution to the dose
603 covariance matrix); indeed, this error on the conversion factor will vary from one site to another
604 depending on the concentrations in K, U and Th (which are generally unknown if a counting technique
605 is used), and even within one site from one sample to another (again by unknown amounts since the
606 variability in K, U and Th is unknown).

607   The data acquired with field gamma spectrometers may be analysed in two ways: the 'window'
608 technique (see, *e.g.*, Aitken, 1985) corresponds to classical spectrometry analysis; in this case, the
609 structure of uncertainties is the same as that for beta dose rates determined from high-resolution
610 gamma spectrometry (Eq. 11~~3~~). On the other hand, threshold techniques consist of taking advantage
611 of proportionality between gamma dose rates and (i) the number of counts recorded per unit time
612 above a threshold (Løvborg and Kirkegaard, 1974) or (ii) the energy deposited per unit time above a
613 threshold (energy threshold: Guérin and Mercier, 2011; Miallier *et al.*, 2009). In the former case, the
614 conversion from count rate to dose rate depends on the emitting radioelement, so no systematic error
615 term may be isolated. Conversely, in the latter case (energy threshold), this dependency is negligible
616 (Guérin and Mercier, 2011). As a result, the error on the dose rate of the calibration standard may be
617 considered as systematic, and thus contribute one term in the non-diagonal elements of the covariance
618 matrix. Difficulties in implementing the threshold technique may occur in low dose rates
619 environments, because energy calibration may not be straightforward if the probe is not self-stabilised.
620 In such cases, routines such as those implemented in the R package 'gamma' (Lebrun et al., 2020) may
621 be put to profit. Finally, ageing of the crystal may also result in time-dependent errors – the latter must
622 be taken care of by regular calibration experiments.

623   **7.2. Error terms neglected in this study**

624   As mentioned earlier in the section devoted to dose rate uncertainties, there are many
625 possibilities to quantify, but also to consider errors on dose rate measurements; one could mention
626 here the uncertainties on attenuation factors and water correction factors. However, both of these
627 factors are dependent on the infinite matrix assumption: attenuation in grains implies that something
628 other than the grains does not attenuate radiation (i.e. conservation of energy implies that if there is
629 lower dose rate inside a quartz grain, there must be a higher dose rate somewhere else – cf. Guérin *et*
630 *al.*, 2012a) ; water correction factors are often calculated assuming a homogeneous mixture of water
631 and other sedimentary components (Zimmerman, 1971; Aitken and Xie, 1990; note: the composition
632 of the sediment also necessarily affects the ratios of electron stopping powers and photon interaction
633 cross-sections – see Nathan and Mauz, 2008, for a discussion). Limitations of this infinite matrix
634 assumption, which is not met in sand samples at the scale of beta dose rates, have already been
635 pointed out (Guérin and Mercier, 2012; Guérin *et al.*, 2012a; Martin *et al.,* 2015). Consequently, it
636 seems that routine determination of a realistic standard deviation of the attenuation and water
637 content correction parameters is not straightforward.

638   Dose rate conversion factors were assumed above to be known without error; however,
639 estimation errors affect half-lives, emission probabilities, average emitted energies, etc. Liritzis *et al.*
640 (2013) took these uncertainties into account to estimate standard deviations of the dose rate
641 conversion factors (in practice, these standard deviations amount to ~1% for K dose rates, ~2% for U
642 and ~2% for Th). These standard deviations could be included as sources of systematic errors when the
643 contributions of K, U and Th are determined separately (note: when this is not the case, as when
644 dosimeters are used for gamma dose rate estimation, or when beta counting is implemented for beta
645 dose rate assessment, these sources of errors should be treated as random).

646   In this study, we worked with coarse grain quartz extracts that had been etched with HF to
647   remove the alpha-irradiated part of the grains. This being said, if alpha dose rates are taken into
648   account, then the situation becomes similar to that of beta dose rates treated above; however, the
649   sensitivity to alpha irradiation must then be taken into account. It is rather frequent in such a case to
650   use published values from the literature (e.g., Tribolo *et al.*, 2001; Mauz *et al.*, 2006). Depending on
651   the geological origin of the quartz (one or more sources), one may then assume either systematic or
652   random errors on the alpha sensitivity.

653   **7.3. Publication habits and re-analysis of previously published ages**

654   Compared to other statistical models for OSL dating, the Bayesian models implemented in
655   'BayLum' appear rather complicated, at least partly because modelling starts from the measured OSL
656   data. By comparison, the input data to the CDM or the Average Dose Model (ADM: Guérin *et al.*, 2017)
657   are lists of equivalent doses and associated uncertainties, which means that OSL measurements have
658   already been analysed to derive equivalent doses. Combès *et al.* (2015) argued that their complete
659   model (implemented in 'BayLum'), relating all the variables to one another, produces a more
660   homogeneous and consistent inference compared to consecutive inferences (and indeed, when
661   approaching saturation, *i.e.* when equivalent doses and associated uncertainties can hardly be
662   parameterised, Heydari and Guérin (2018) demonstrated the advantage of 'BayLum' models compared
663   to parametric models such as the CDM and ADM). However, working with lists of equivalent doses and
664   uncertainties – or even with estimates of central doses and associated uncertainties – taken as
665   observations would make the Bayesian modelling proposed in 'BayLum' and described in this paper
666   more straightforward and transparent. Such an approach, called the 'two-steps' model by Combès and
667   Philippe (2017; see also Millard 2006a, 2006b, for earlier, similar models), would also offer the
668   advantage of allowing re-analysis of already published data to derive more precise chronologies.
669   However, for this purpose, the breakdown of all uncertainties and related standard deviations of errors
670   is needed; nowadays, providing such key information for the modelling is not in the luminescence
671   dating community's publication habits. That being said, with the growing number of meta-analyses of
672   previously published data, and the availability to use models such as 'BayLum' to combine
673   measurements with systematic errors, these habits might evolve in the future.

674   **7.4. Notes of caution**

675   As always when working with statistical models, one should first and foremost evaluate the
676   measured data in the light of sampling context. We already mentioned the importance of grain
677   selection (section 2.2.); but, perhaps more importantly, and especially since users of 'BayLum' have to
678   make modelling choices (*e.g.*, regarding the dose-response curves fitted to OSL measurements or the
679   distribution of individual equivalent doses around the central dose), it is crucial to carefully examine
680   data and assess their quality before building potentially sophisticated models.

681   We would like to emphasise a few warnings regarding modelling samples in stratigraphic
682   constraints, and the association of ages obtained by different methods. We would advise users, before
683   combining, *e.g.* radiocarbon and OSL ages, first to thoroughly examine the corresponding datasets
684   independently: how were the data produced (with which experimental procedure)? Are the provided
685   uncertainties reliable (or is there an unrecognised source of error that should be included in the
686   evaluation of uncertainty)? Users are also encouraged to examine the consistency of results produced
687   by each method, in light of the stratigraphy. In a second stage, before modelling of independent ages,
688   we would recommend assessing the consistency of these datasets – do they (at least broadly) agree?
689   If not, can a parsimonious explanation be found? For example, it is rather common, when performing
690   Bayesian modelling with tools such as *OxCal*, to observe a large fraction of ages considered as outliers;

691   such observations should urge users to examine their data again and come up with likely explanations
692   (note: to this date, no outlier model has been developed for the OSL ages in 'BayLum').

693   When it comes to imposing ordering constraints between ages as a result of stratigraphic
694   observations, it is, of course, essential to leave no doubt about the validity of these stratigraphic
695   constraints (the results of a model depend on the assumptions that are made, and the order in ages is
696   a very strong constraint). Perhaps more importantly, even when stratigraphic constraints are valid, it
697   is possible that applying them will not improve the statistical inference.

698   A simple example to illustrate this point is that of two superimposed, distinct layers (so that a
699   stratigraphic order is clear) whose true ages are equal (or in practice, for which the age difference is
700   negligible compared to the typical uncertainties of the implemented dating method). In such a case,
701   modelling the ages with stratigraphic constraints is likely to result in a loss of accuracy (the age of the
702   older layer will be overestimated, and that of the younger layer underestimated) compared to a model
703   where no stratigraphic constraints are imposed. Future developments of the 'BayLum' package might
704   include the possibility to test different modelling scenarios by comparing the agreement between the
705   observations and the posterior probability densities, for example using the Bayes Information Criterion
706   (BIC).

707   **8. Conclusion**

708   New models for building chronologies based on OSL, with the possibility to incorporate
709   radiocarbon, have been proposed in the literature (Combès *et al.*, 2015; Combès and Philippe, 2017).
710   These models have been demonstrated to improve the chronological inference based on OSL data and
711   in particular, the accuracy of OSL ages (Guérin *et al.*, 2015; Heydari and Guérin, 2018). The R package
712   'BayLum' was developed to implement these models; Lahaye *et al.* (2018), Carter *et al.* (2019), Heydari
713   *et al.* (2020, 2021) have used some of them to establish the chronologies of sedimentary sequences
714   dated by OSL, resulting in generally more precise chronologies.

715   In this article, we have presented a case study on how to build simple models and observe
716   output data, particularly through bivariate plots of age probability densities. Then, we have shown how
717   to include stratigraphic constraints in the models; we have described how to fill the covariance
718   matrices to account for systematic errors in OSL age estimation; and we have shown the effect of
719   including independent age information in the models, namely radiocarbon ages. Different tools to
720   visualise and further analyse the output of 'BayLum' were demonstrated.

721   As a result, it is now possible to use various information often available in practice when dating
722   stratigraphic sequences. Age inferences based on OSL and independent data (*e.g.,* radiocarbon) in
723   stratigraphic constraints are expected to gain in accuracy, precision and robustness, through the
724   application of such Bayesian models.

725

726   **Acknowledgements**

727   The authors thank Andrew Millard (twice), Eddie Rhodes, Rex Galbraith and two anonymous referees
728   for helpful comments on previous versions of this article. This study received financial supports of the
729   Région Aquitaine (in particular through the CHROQUI programme) and of the LaScArBx project (project
730   n∘ ANR-10-LABX-52). M. Autzen, and J.P. Buylaert received funding from the European Research
731   Council (ERC) under the European Union's Horizon 2020 research and innovation programme ERC-
732   2014-StG 639904-RELOS. G. Guérin received funding from the European Research Council (ERC) under

733 the European Union's Horizon 2020 research and innovation programme (Grant Agreement 851793)
734 Project QuinaWorld (ERC-StG-2019).

735 **References**

736 Aitken, M. J., 1985. Thermoluminescence dating. Academic Press, London, 359 p.

737 Aitken M.J., 1998. An introduction to optical dating. Oxford University press, Oxford, 267 p.

738 Aitken, M.J. and Xie, J., 1990. Moisture correction for annual gamma dose, Ancient TL 8 (2), pp. 6–9.

739 Ankjærgaard, C., Murray, A.S., 2007. Total beta and gamma dose rates in trapped charge dating based
740 on beta counting. Radiation Measurements, 42, 352-359.

741 Bronk Ramsey, C., Lee, S., 2013. Recent and Planned Developments of the Program OxCal.
742 Radiocarbon, 55(2-3), 720-730.

743 Buck CE, Kenworthy JB, Litton CD, Smith AFM. 1991. Combining archaeological and radiocarbon
744 information: a Bayesian approach to calibration. Antiquity 65(249):808–21.

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

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

751 Chevrier, .B., Lespez, L., Lebrun B., Garnier, A., Tribolo, C., Rasse, M., Guérin, G., Mercier, N., Camara,
752 A., Ndiaye, M., Huysecom, E. New data on settlement and environment at the 2 Pleistocene/Holocene
753 boundary in Sudano-Sahelian West Africa: 3 interdisciplinary investigation at Fatandi V, Eastern
754 Senegal. PlosOne, accepted for publication.

755 Christophe, C., Philippe, A., Kreutzer, S., Guérin, S., 2020: 'BayLum': Chronological Bayesian Models
756 Integrating Optically Stimulated Luminescence and Radiocarbon Age Dating. R package, version 0.2.0.
757 https://CRAN.R-project.org/package='BayLum'

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

761 Combès, B. and Philippe, A., 2017. Bayesian analysis of individual and systematic multiplicative errors
762 for estimating ages with stratigraphic constraints in optically stimulated luminescence dating,
763 Quaternary Geochronology, 39, 24—34.

764 Combès, B., Lanos, P., Philippe, A., Mercier, N., Tribolo, C., Guérin, G., Guibert, P., Lahaye, C., 2015. A
765 Bayesian central equivalent dose model for optically stimulated luminescence dating. Quaternary
766 Geochronology 28, 62-70.

767 Cunningham, A.C., Murray, A.S., Armitage, S.J., Autzen, M., 2018. High-precision natural dose rate
768 estimates through beta counting. Radiation Measurements, in press.

769 Duller, G.A.T., 2015. The Analyst software package for luminescence data: overview and recent
770 improvements. Ancient TL 33, 35-42.

771 Duller, G.A.T., Bøtter-Jensen, L., Murray, A.S., Truscott, A.J., 1999. Single grain laser luminescence
772 (SGLL) measurements using a novel automated reader. Nuclear Instruments and Methods B 155, 506–
773 514.

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

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

779 Gelman, A., Rubin, D., 1992. Inference from Iterative Simulation using Multiple Sequences. Statistical
780 Science 7, 457–511.

781 Guérin, G., Mercier, N., 2011. Determining gamma dose rates by field gamma spectroscopy in
782 sedimentary media: results of Monte Carlo simulations. Radiation Measurements, 46 (2), 190-195.

783 Guérin, G., Mercier, N., 2012. Preliminary insight into dose deposition processes in sedimentary media
784 on a grain scale: Monte Carlo modelling of the effect of water on gamma dose-rates. Radiation
785 Measurements, 47, 541-547.

786 Guérin, G., Mercier, N. and Adamiec, G, 2011. Dose-rate conversion factors: update. Ancient TL 29, 5-
787 8.

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

790 Guérin, G., Discamps, E., Lahaye, C., Mercier, N., Guibert, P., Turq, A., Dibble, H., McPherron, S.,
791 Sandgathe, D., Goldberg, P., Jain, M., Thomsen, K., Patou-Mathis, M., Castel, J.-C., Soulier, M.-C.,
792 2012b. Multi-method (TL and OSL), multi-material (quartz and flint) dating of the Mousterian site of
793 the Roc de Marsal (Dordogne, France): correlating Neanderthals occupations with the climatic
794 variability of MIS 5-3. Journal of Archaeological Science, 39, 3071-3084.

795 Guérin, G., Combès, B., Lahaye, C., Thomsen, K. J., Tribolo, C., Urbanova, P., Guibert, P., Mercier, N.,
796 Valladas, H., 2015a. Testing the accuracy of a single grain OSL Bayesian central dose model with known-
797 age samples. Radiation Measurements 81, 62-70.

798 Guérin, G., Frouin, M., Talamo, S., Aldeias, V., Bruxelles, L., Chiotti, L., Dibble, H. L., Goldberg, P., Hublin,
799 J.-J., Jain, M., Lahaye, C., Madelaine, S., Maureille, B, McPherron, S. P., Mercier, N., Murray, A. S.,
800 Sandgathe, D., Steele, T. E., Thomsen, K. J., Turq, A., 2015b. A Multi-method Luminescence Dating of
801 the Palaeolithic Sequence of La Ferrassie Based on New Excavations Adjacent to the La Ferrassie 1 and
802 2 Skeletons, au Journal of Archaeological Science 58, 147-166.

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

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

810 Guibert, P., 2002. Progrès récents et perspectives. Habilitation à diriger des recherches, 10 juillet 2002.
811 Datation par thermoluminescence des archéomatériaux : recherches méthodologiques et appliquées
812 en archéologie médiévale et en archéologie préhistorique, 3. Université de Bordeaux, pp. 27-50.

Guibert P.et Schvoerer M., 1991 - TL dating : Low background gamma spectrometry as a tool for the determination of the annual dose, Nuclear Tracks and Radiation Measurements 18, 231-238.

Guibert P., Schvoerer M., Etcheverry M.P., Szepertyski B. et Ney C., 1994. IXth millenium B.C. ceramics from Niger : detection of a U-series disequilibrium and TL dating, Quaternary Science Reviews, 13, 555-561.

Guibert P., Lahaye C., Bechtel F., 2009. The importance of U-series disequilibrium of sediments in luminescence dating: A case study at the Roc de Marsal Cave (Dordogne, France), Radiation Measurements 44, 223-231.

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

Heydari, M., Guérin, G., 2018. OSL signal saturation and dose rate variability: investigating the behaviour of different statistical models. Radiation measurements 120, 96-103.

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

Heydari, M., Guérin, G., Zeidi, M., 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, 1–15. doi:10.1016/j.jhevol.2020.102926

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., 2008a. Ages for the Middle Stone Age of southern Africa: implications for human behavior and dispersal. Science 322, 733-735.

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.

Kreutzer, S., Dietze, M., Burow, C., Fuchs, M. C., Schmidt, C., Fischer, M., Friedrich, J., Mercier, N., Smedley, R. K., Christophe, C., Zink, A., Durcan, J., King, G.E.,Phlippe, A., Guérin, G., Riedesel, S., Autzen, M., Guibert, P.,.2020. Luminescence: Comprehensive Luminescence Dating Data Analysis. R package, version 0.9.7. https://CRAN.R-project.org/package=Luminescence.

Kreutzer, S., Martin, L., Guérin, G., Tribolo, C., Selva, P., Mercier, N., 2018. Environmental Dose Rate Determination Using a Passive Dosimeter: Techniques and Workflow for alpha-Al2O3:C Chips. Geochronometria 45, 56–67.

Lahaye, C., Guibert, P., Bechtel, F., 2012, Uranium series disequilibrium detection and annual dose determination: A case study on Magdalenian ferruginous heated sandstones (La Honteyre, France), Radiation Measurements, 47, 786-789.

Lahaye, C., Guérin, G., Gluchy, M., Hatté, C., Fontugne, M., Clemente-Conte, I., Santos, J.C., Villagran, X., Da Costa, A., Borges, C., Guidon, N. and Boëda, E., 2018. Another site, same old song: the Pleistocene-Holocene archaeological sequence of Toca da Janela da Barra do Antonião-North, Piauí, Brazil. Quaternary Geochronology, in press.

Lanos, P., Philippe, A., 2018, Event date model: a robust Bayesian tool for chronology building, Communications for Statistical Applications and Methods, 25, 1–28.

Code de champ modifié

854 Lebrun, B., Frerebeau, N., Paradol, G., Guérin, G., Mercier, N., Tribolo, C., Lahaye, C., Rizza, M., 2020.
855 gamma: An R Package for Dose Rate Estimation from In-Situ Gamma-Ray Spectrometry Measurements.
856 Ancient TL 18, 1-5.

857 Liritzis, I., Stamoulis, K., Papachristodoulou, C., Ioannides, K., 2013. A re-evaluation of radiation dose-
858 rate conversion factors. Mediterranean Archaeology and Archaeometry, 13, 1-15.

859 Løvborg, L., Kirkegaard, P., 1974. Response of 3" x 3" NaI(Tl) detec-tors to terrestrial gamma radiation.
860 Nuclear Instruments and Methods 121, 239-251.Mauz, B., Packman, S. & Lang, A. 2006: The alpha
861 effectiveness in silt-sized quartz: New data obtained by single and multiple aliquot protocols. Ancient
862 TL 24, 47–52.

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

864 Mercier, N. and Falguères, C., 2007. Field gamma dose rate measurement with a NaI (Tl) detector: re-
865 evaluation of the "threshold" technique. Ancient TL, 25, 1-4.

866 Mercier, N., Kreutzer, S., Christophe, C., Guérin, G., Guibert, P., Lahaye, C., Lanos, P., Philippe, A.,
867 Tribolo, C., 2016. Bayesian statistics in luminescence dating: The 'baSAR'-model and its
868 implementation in the R package 'Luminescence'. Ancient TL 34, 14–21.

869 Miallier, D., Guérin, G., Mercier, N., Pilleyre, T., Sanzelle, S., 2009. The Clermont radiometric reference
870 rocks: a convenient tool for dosimetric purposes. Ancient TL, 27 (2), 37-42.

871 Millard, A., 2006a. Bayesian analysis of ESR dates, with application to border cave. Quaternary
872 Geochronology 1, 159-166.

873 Millard, A., 2006b. Bayesian analysis of Pleistocene chronometric methods. Archaeometry 48, 359-
874 375.

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

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

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

881 Murray, A.S., Buylaert, J.-P., Thiel, C., 2015. A luminescence dating intercomparison based on a Danish
882 Beach-ridge sand. Radiation Measurements 81, 32-38.

883 Nathan R.P, Mauz B., 2008. On the dose-rate estimate of carbonate-rich sediments for trapped charge
884 dating. Radiation Measurements 43, 14-25.

885 Nelson, M. S., Rittenour, T. M., 2015. Using grain-size characteristics to model soil water content:
886 Application to dose-rate calculation for luminescence dating. Radiation Measurements 81, 142-149.

887 Philippe, A., Guérin, G., Kreutzer, S., 2019. ''BayLum'' an R package for Bayesian Analysis of OSL Ages
888 & Chronological Modelling. Quaternary Geochronology 49, 16-24.

889 Philippe, A., Vibet, M., 2020. "Analysis of Archaeological Phases Using the R Package ArchaeoPhases.
890 Journal of Statistical Software, Code Snippets, 93(1), 1–25.

891 Prescott, J.R., Hutton, J.T., 1988. Cosmic ray and gamma ray dosimetry for TL and ESR. Nuclear Tracks
892 and Radiation Measurements, 14, 223-227.

893 R Core Team, 2020. R: A Language and Environment for Statistical Computing. R Foundation for
894 Statistical Computing, Vienna, Austria. https://r-project.org

895 Reimer, P. J., Austin, W. E., Bard, E., Bayliss, A., Blackwell, P. G., Ramsey, C. B., ... & Grootes, P. M.
896 (2020). The IntCal20 northern hemisphere radiocarbon age calibration curve (0–55 cal
897 kBP). Radiocarbon, 62(4), 725-757.

898 Rhodes, E. J., Ramsey, C. B., Outram, Z., Batt, C., Willis, L., Dockrill, S., & Bond, J. (2003). Bayesian
899 methods applied to the interpretation of multiple OSL dates: high precision sediment ages from Old
900 Scatness Broch excavations, Shetland Isles. Quaternary Science Reviews, 22(10-13), 1231-1244.

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

903 Thomsen, K.J., Murray, A.S., Bøtter-Jensen, L., 2005. Sources of variability in OSL dose measurements
904 using single grains of quartz. Radiation Measurements, 39, 47-61.

905 Thomsen, K.J., Murray, A.S., Buylaert, J.-P., Jain, M., Helt-Hansen, J., Aubry, T., 2016. Testing single-
906 grain quartz OSL methods using known age samples from the Bordes-Fitte rockshelter (Roches d'Abilly
907 site, Central France). Quaternary Geochronology 31, 77-96.

908 Tribolo, C., Mercier, N., Valladas, H., 2001. Alpha sensitivity determination in quartzite using an OSL
909 single aliquot procedure. Ancient TL 19, 47–50.

910 Wintle, A.G., Murray, A.S., 2006. A review of quartz optical stimulated luminescence characteristics
911 and their relevance in single-aliquot regeneration dating protocols. Radiation Measurements, 41, 369-
912 391.

913 Zimmerman, D.W., 1971. Thermoluminescence dating using fine grains from pottery, Archaeometry
914 13, pp. 29–52.

915

Code de champ modifié

Mis en forme : Anglais (Royaume-Uni)

 Figures

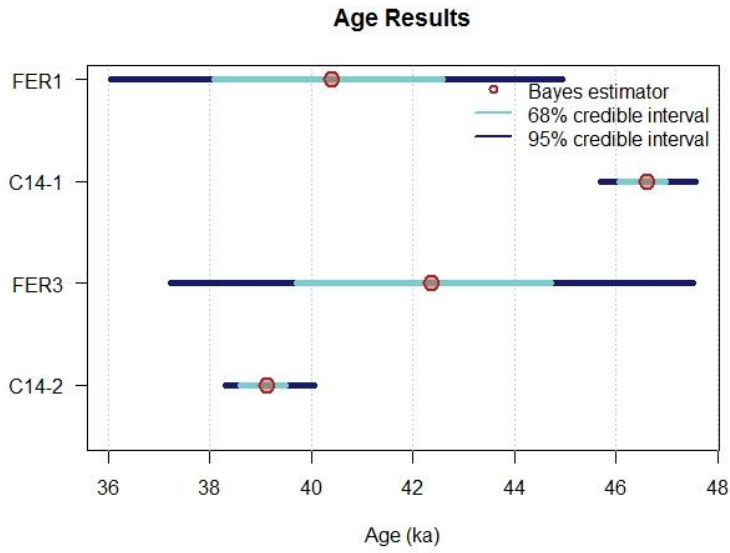



Fig. 1: Age estimates for OSL samples FER 1 and FER 3. The red circles indicate the Bayes estimates of the age (*i.e.* the most likely values) for each sample; the cyan and blue bars represent the 68% and 95% credible intervals, respectively. For the two radiocarbon ages (C14-1 and C14-2), the reader is refereed to section 6.

924

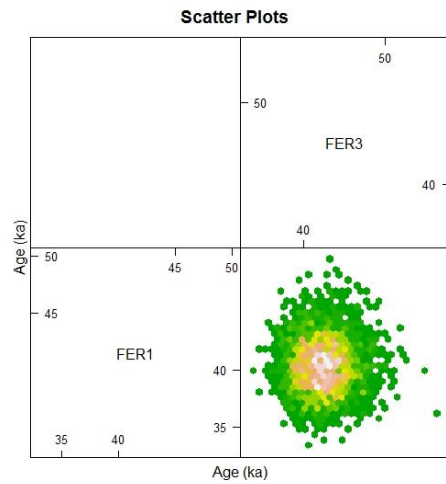



Fig. 2: Bivariate scatter plot as hexagon plot presentation of a sample of observations from the joint
posterior distribution of the two OSL ages considered independently (no stratigraphic constraints, no
off-diagonal members in the covariance matrix). In such a plot, each point corresponds to one
realisation of the ages of the two samples generated by the MCMC. Note: the reason for having this
figure in the cell of an array is not visible here; it becomes useful when calculating ages for more than
2 samples, in which case for each pair of samples, a similar plot appears in the appropriate cell.

931

932

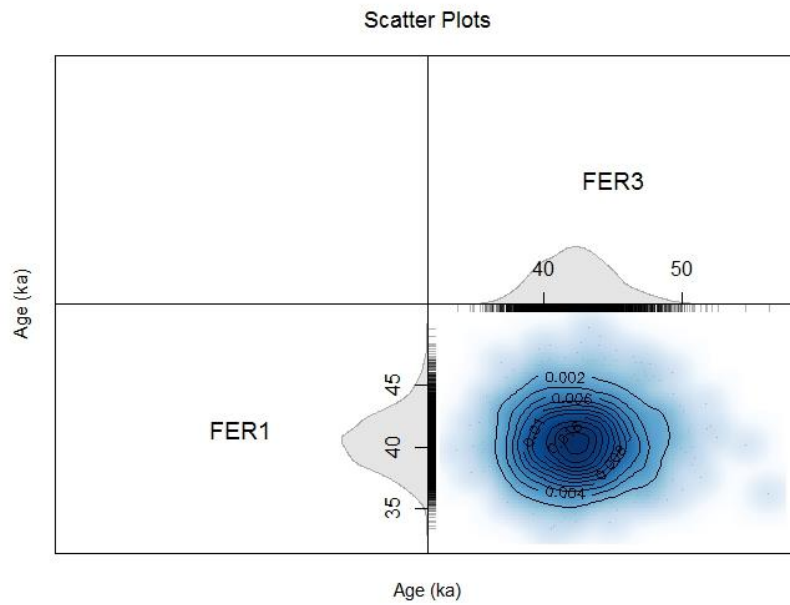
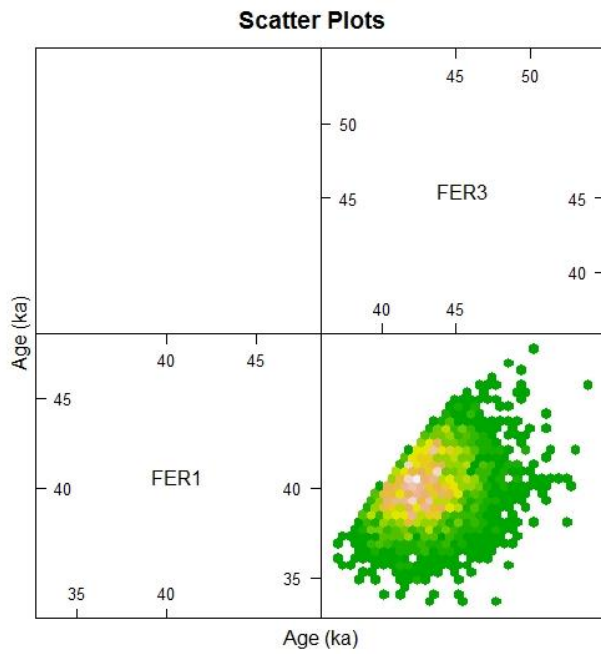



Fig. 3: Probability densities for the OSL ages estimated jointly with the same model as that used to generate Fig. 2, based on Kernel Density Estimates (KDE), and marginal probability densities. The bell-shape and symmetry of the scatter plot indicate the absence of correlation between the two ages.

936

937
Fig. 4: Bivariate scatter plot from the joint posterior distribution of the ages of samples FER 1 and FER
3 when a stratigraphic constraint is applied (sample FER 1 is younger than sample FER 3) but with no
off-diagonal members in the covariance matrix. The truncation in the upper-left hand corner scatter
plot indicates the effect of the stratigraphic constraint.

942

943

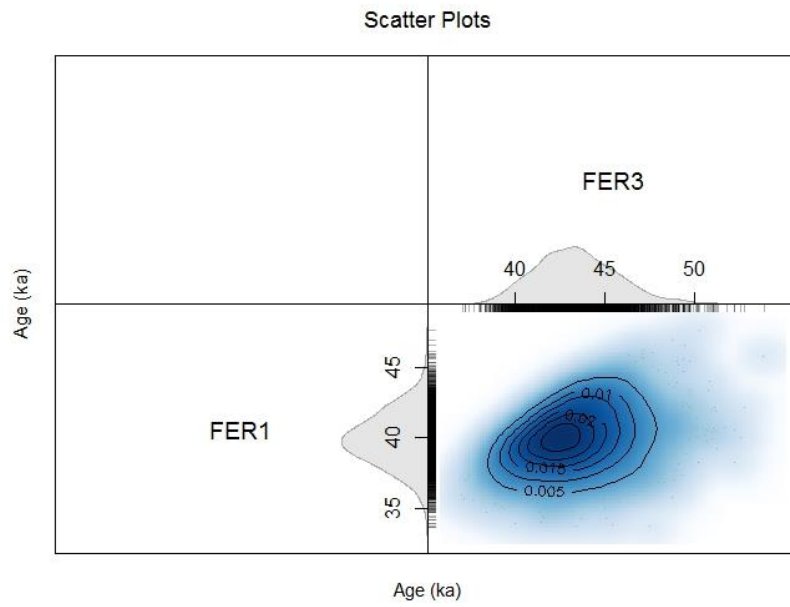
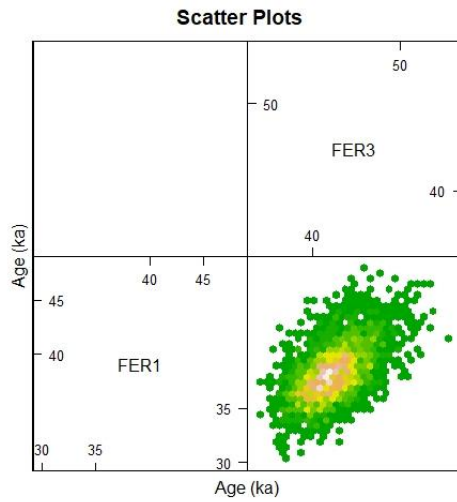
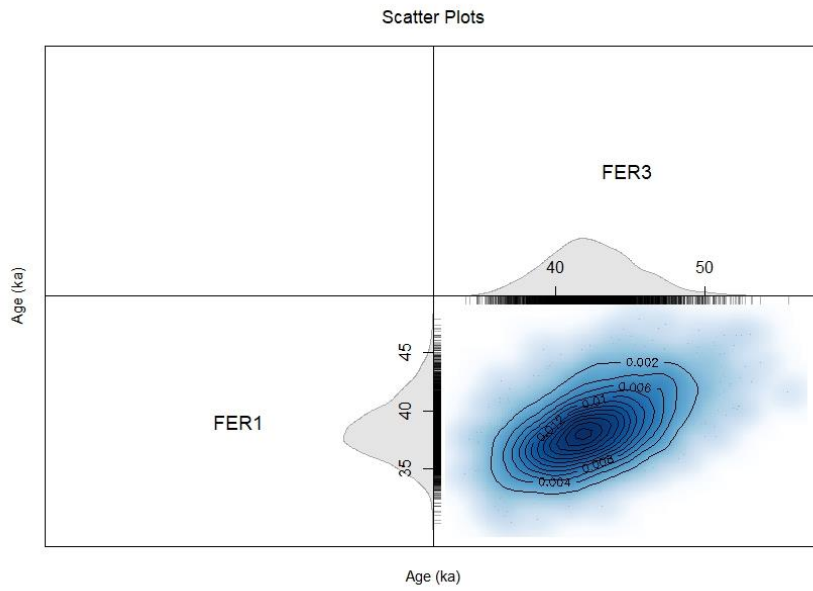



Fig. 5: Probability densities for the OSL ages estimated jointly, using the same model as that
implemented to generate Fig. 4 (stratigraphic constraint, no covariance matrix).

946

947
948 **Fig. 6:** Bivariate scatter plot from the joint posterior distribution of the ages of samples FER 1 and FER
949 3 when a stratigraphic constraint is applied (sample FER 1 is younger than sample FER 3) and off-
950 diagonal members of covariance matrix are used to model systematic errors (note: in this case, for
951 illustrative purposes we used a simplistic covariance matrix – see section 5.3.1. for details). The
952 truncation in the upper-left hand corner scatter plot indicates the effect of the stratigraphic constraint.

953

954
955 **Fig. 7:** Probability densities for the OSL ages estimated jointly, using the same model as that
956 implemented to generate Fig. 6 (stratigraphic constraint and off-diagonal members in the covariance
957 matrix). The positive correlation in the joint posterior density reflects the effect of modelling the
958 systematic errors with a covariance matrix (and, to some degree, of the stratigraphic constraint).

959

960 **Table 1.** Summary of Credible Intervals for the ages (in ka) of samples FER 1 and FER 3 estimated in
961 the different modelled scenarios.

Sample	68% Credible Interval		95% Credible Interval	
	lower	upper	lower	upper
Independent				
FER 1	36.0	40.5	34.1	43.3
FER 3	38.9	44.6	36.6	47.8
In stratigraphy				
FER 1	36.2	40.4	34.3	42.9
FER 3	40.0	45.0	38.1	48.5
No stratigraphic constraint, with 'simplistic' covariance (section 5.3.1)				
FER 1	36.0	40.8	33.9	43.8
FER 3	39.2	45.4	36.7	48.1
In stratigraphy, with realistic covariance (section 5.3.2)				
FER 1	36.1	40.5	34.2	42.6
FER 3	39.8	45.3	37.8	48.6
In stratigraphy, with covariance and a 'young' radiocarbon age				
FER 1	35.2	39.4	33.3	41.2
FER 3	39.2	42.2	36.9	42.3
In stratigraphy, with covariance and an 'old' radiocarbon age				
FER 1	38.7	43.5	36.2	46.2
FER 3	46.1	48.7	46.1	51.5

962

963

964 **Table 2.** List of physical units and associated uncertainties used in this work. The letter i in subscript
965 indicates a sample specific value, its absence a common value shared between samples. The letter s
966 indicates absolute uncertainties, while σ is used for relative uncertainties.

Physical unit	Notation	Systematic uncertainty	Random uncertainty
Laboratory source dose rate	$\dot{d}_{lab}$	σ_{lab}	
Cosmic dose rate	$\dot{d}_{cosmic,i}$		$s_{cosmic,i}$
K concentration	$[K]_i$	σ_K	$\sigma_{K,i}$
U concentration	$[U]_i$	σ_U	$\sigma_{U,i}$
Th concentration	$[Th]_i$	σ_{Th}	$\sigma_{Th,i}$
Internal dose rate	$\dot{d}_{int}$	s_{int}	
Gamma dose rate	$\dot{d}_{\gamma,i}$	σ_γ	$\sigma_{\gamma,i}$
Water content	WF_i		$s_{WF,i}$

967