Conference Proceedings of the 5ᵗʰ Asia Pacific Luminescence and Electron Spin Resonance Dating Conference
October 15ᵗʰ-17ᵗʰ, 2018, Beijing, China

Guest Editor: Grzegorz Adamiec

ANALYZING STATISTICAL AGE MODELS TO DETERMINE THE EQUIVALENT DOSE AND BURIAL AGE USING A MARKOV CHAIN MONTE CARLO METHOD

JUN PENG

School of Resource, Environment and Safety Engineering, Hunan University of Science and Technology,
Hunan, Xiangtan 411201, China

Abstract: In optically stimulated luminescence (OSL) dating, statistical age models for equivalent dose (D_e) distributions are routinely estimated using the maximum likelihood estimation (MLE) method. In this study, a Markov chain Monte Carlo (MCMC) method was used to analyze statistical age models, including the central age model (CAM), the minimum age model (MAM), the maximum age model (MXAM), *etc.* This method was first used to obtain sampling distributions on parameters of interest in an age model using D_e distributions from individual sedimentary samples and subsequently extended to simultaneously extract age estimates from multiple samples with stratigraphic constraints. The MCMC method allows for the use of Bayesian inference to refine chronological sequences from multiple samples, including both fully and partially bleached OSL dates. This study designed easily implemented open-source numeric programs to perform MCMC sampling. Measured and simulated D_e distributions are used to validate the reliability of dose (age) estimates obtained by this method. Findings from this study demonstrate that estimates obtained by the MCMC method can be used to informatively compare results obtained by the MLE method. The application of statistical age models to multiple OSL dates with stratigraphic orders using the MCMC method may significantly improve both the precision and accuracy of burial ages.

Keywords: OSL dating, statistical age models, equivalent dose, MCMC sampling.

1. INTRODUCTION

In optically stimulated luminescence (OSL) dating, the assessment of certain parameters of interest (*i.e.*, D_e) is essential in estimating sample age. Routinely, a large set of aliquots (grains) are analyzed from which a single D_e (*i.e.*, the characteristic D_e determined using a statistical model) is ultimately obtained along with its standard error estimate. Galbraith and Roberts (2012) reviewed the most commonly used statistical age models in determining the characteristic D_e, namely, from the simple central age model (CAM) (Galbraith *et al.*, 1999) to more sophisticated ones, such as the minimum age model (MAM) (Galbraith *et al.*, 1999).

Corresponding author: J. Peng
e-mail: pengjun10@mails.ucas.ac.cn

Maximum likelihood estimation (MLE), which provides point estimates of parameters and associated standard errors, is routinely adopted to estimate parameters used in these statistical age models by maximizing the logged-likelihood function (Galbraith *et al.*, 1999; Galbraith and Roberts, 2012). The burial age is obtained by dividing the characteristic D_e estimated using MLE by the average environmental dose rate of the sample, and the associated standard error is estimated using propagation of error formulas (referred to as the standard approach). However, the standard approach only provides a coarse characterization of the statistics associated with burial age. Moreover, it does not permit the inclusion of additional information, such as order constraints between the ages of a given stratigraphic sequence to improve age estimates (Combès and Philippe, 2017).

In contrast, the Markov chain Monte Carlo (MCMC) methods construct a Markov chain while converging to the desired sampling distribution by starting from random initials and seeking to characterize posterior distributions for parameters of interest (Gelman *et al.*, 2014). Furthermore, MCMC methods treat model parameters as full probability distributions and produce samples from the joint posterior density of parameters that are then summarized for the purpose of Bayesian inference. By doing so, the uncertainty over the range of potential parameter values is also estimated, rather than only a single point estimate (Annis *et al.*, 2017). Summary statistics (such as the mean, the median, the standard deviation, the confidence interval, *etc.*) are obtained directly from randomly generated samples.

Such MCMC methods have increasingly been employed to analyze OSL data for Bayesian inferences (e.g., Millard, 2004; Sivia *et al.*, 2004; Huntriss, 2008; Zink, 2013, 2015; Peng and Dong, 2014; Combès *et al.*, 2015; Cunningham *et al.*, 2015a, 2015b; Guérin *et al.*, 2015; Mercier *et al.*, 2016; Peng *et al.*, 2016a; Christophe *et al.*, 2018; Philippe *et al.*, 2019). For example, Sivia *et al.* (2004) used an MCMC method to analyze the finite mixture age mode (FMM) (Galbraith and Green, 1990; Roberts *et al.*, 2000) as well as a noised FMM using unlogged D_e sets. Cunningham *et al.* (2015a, 2015b) developed an age model to account for effects of variation in aliquot size and luminescence sensitivity on partially-bleached D_e distributions while using an MCMC method to obtain the posterior of parameter estimates. Combès *et al.* (2015) proposed a hierarchical model to describe the central D_e for fully bleached aliquots (grains). A Metropolis-within-Gibbs sampler is used to obtain the posterior of the characteristic dose. The model has been further expanded to analyse poorly bleached samples using a Gaussian mixture model (e.g., Christophe *et al.,* 2018). Peng *et al.* (2016a) used a Slice-within-Gibbs sampling algorithm to obtain the posterior distribution of parameters in statistical age models reviewed by Galbraith and Roberts (2012). However, this model did not take into account the posterior distribution for burial age.

The first objective of this study is to estimate statistical model parameters (such as the CAM, the MAM, *etc.*) and obtain the posterior distribution for burial age through MCMC sampling under a Bayesian framework, using a combination of aliquot-level (grain-level) dose and sample-level dose-rate datasets. The burial dose and age estimated by MCMC are then compared with those estimated using the standard approach (*i.e.*, MLE).

Bayesian models utilize stratigraphic relationships between samples to increase the precision of dates and to provide a sophisticated protocol to combine multiple age estimates into a meaningful chronological framework (Ramsey, 1995, 2008). By incorporating the stratigraphic principle that shallower sediments in the stratum sequence are younger than deeper sediments into age-depth models, Bayesian inference can reduce the uncertainty of dates wherein age probability distributions overlap. Age-depth modelling using Bayesian inference was initially developed to analyze radiocarbon dating data (Ramsey, 1995; Bayliss and Ramsey, 2004). Over the past decade, the application of Bayesian inference to model depth-age relationships using luminescence data have also gained widespread usage (e.g., Rhodes *et al.*, 2003; Cunningham and Wallinga, 2012; Lanos and Philippe, 2015; Combès and Philippe, 2017; Zeeden *et al.*, 2018; Tamura *et al.*, 2019).

During the routine application of statistical age models, such as the CAM and MAM, characteristic D_e values of individual samples are estimated separately (e.g., Schmidt *et al.*, 2012; Arnold *et al.*, 2009), even when samples are collected from different depths in the same profile sequence (e.g., Arnold *et al.*, 2007; Kunz *et al.*, 2013; Peng *et al.*, 2016b) and when age-depth relationship between samples may have been used as additional information to improve date estimates. Brill *et al.* (2015) applied statistical age models (including CAM and MAM) to obtain luminescence age estimates (and associated standard errors) for samples collected from a number of depositional sequences, and they used the age-depth modelling program OxCal to improve the accuracy and precision of these ages by making use of their relative stratigraphic order. However, this two-step protocol that consists of calculating burial ages using a standard approach and further perfect these age estimates using Bayesian inference may fail in characterizing distributions of interest because it decomposes the global inference into a two-step assessment that is likely to lead to a significant loss of information during each step in the process (Combès *et al.*, 2015). Cunningham and Wallinga (2012) employed a bootstrap likelihood protocol to obtain the distribution of the characteristic D_e for a sequence of young fluvial samples using an unlogged version of the MAM. The replicate characteristic D_e values obtained by bootstrapping are used to construct a probability density function for each sample in order to apply Bayesian inference to improve OSL chronologies. However, as reiterated by Cunningham and Wallinga (2012), the boot-

strap likelihood protocol is just an analogue of the likelihood function and does not produce a true likelihood.

The second objective of this study is to apply statistical age models to refine age-depth relationships using D_e datasets from multiple luminescence samples employing an alternative Bayesian inference model. In contrast to the method by Cunningham and Wallinga (2012), in this study characteristic D_e values are sampled directly from the probability density function that describes the statistical model (see the following section for more detail) through MCMC sampling. The joint probability density function for multiple OSL ages is constructed by utilizing a combination of statistical models and measured D_e and dose rate datasets. Both the CAM and MAM are used to flexibly analyze fully and partially bleached OSL dates, respectively, using MCMC in order to combine different types of statistical age models in a mathematically rigorous, geologically reasonable, and philosophically satisfying manner. The performance of the model is then tested on measured and simulated D_e distributions.

2. A BRIEF INTRODUCTION TO GALBRAITH'S STATISTICAL AGE MODELS

The central age model (CAM)

Galbraith *et al.* (1999) recommended using the CAM to provide a representative estimate of the characteristic dose for samples that have not been affected by partial bleaching or sediment mixing. The benefit of the CAM is that it can take any over-dispersion (OD) into account when estimating the weighted mean dose and its standard error (Galbraith and Roberts, 2012). OD provides an estimate of the degree of dispersion that represents the relative spread in the D_e distribution that remains after allowing for measurement uncertainties. Routinely, OD is calculated as a descriptive statistic to reveal sources of variability in D_e (e.g., Thomsen *et al.*, 2005; Jacobs *et al.*, 2006). The OD of D_e distributions in well-bleached analogues is commonly used as an input for statistical age models when analyzing measured sedimentary samples (Galbraith and Roberts, 2012). The likelihood function of the CAM can be described as follows:

$$p(D_j \mid \theta) = CAM(x_j, y_j \mid \mu, \sigma) = \frac{1}{\sqrt{2\pi(x_j^2 + \sigma^2)}} e^{-\frac{(y_j - u)^2}{2(x_j^2 + \sigma^2)}} \quad (2.1)$$

where y_j and x_j represent for the jth logged D_e value and its relative standard error (RSE), respectively. $p(D_j|\theta)$ represents the probability of data D_j given parameter θ. μ and σ represent the parameters of interest (*i.e.*, the logged characteristic D_e and the OD, respectively).

Optimal parameters can be obtained using either an iterative numeric procedure or by direct numerical maximization of the joint-likelihood determined from paired observations $D=\{x_j, y_j \mid j=1,2,3,...,n\}$ (Galbraith *et al.*, 1999). Most free and commercial optimization environments (such as R, Matlab, *etc.*) provide well-designed

optimization routines for optimization usage. The standard errors of parameters are conventionally obtained by inverting the Hessian matrix estimated by finite-difference approximation.

The minimum age model (MAM)

One of the most widely used statistical models for tackling partially-bleached samples is the MAM (Galbraith *et al.*, 1999; Galbraith and Roberts, 2012), which is based on well-established statistical principles. In this model, true logged characteristic D_e value is assumed to be drawn from a truncated Normal distribution, and the lower truncation point (*i.e.*, γ) represents the mean of the logged characteristic dose of fully bleached grains. Two versions of the model are available, namely, the 3-parameter minimum age model (MAM$_3$) and its 4-parameter counterpart (MAM$_4$). The properties of the dataset (*i.e.*, the degree of dispersion, the number of data points, the uncertainties associated with D_e values, *etc.*) may strongly affect the quality of the estimate in the application of the MAM$_4$ (Galbraith and Roberts, 2012). In contrast, the MAM$_3$ usually works well even for lowly dispersed D_e sets, while also being rather numerically stable compared to the MAM$_4$.

Galbraith *et al.* (1999) noted that some datasets, particularly those with small numbers of D_e values or less dispersed distributions, do not warrant fitting with the MAM$_4$, and that a more robust estimate of the characteristic dose may be obtained using the MAM$_3$. Consequently, the MAM$_3$ is conveniently regarded as the "default" model (Arnold *et al.*, 2009) and has been widely applied in the literature to determine burial dose (e.g., Arnold *et al.*, 2007; Schmidt *et al.*, 2012; Kunz *et al.*, 2013). According to Galbraith *et al.* (1999), the likelihood function of the MAM$_3$ can be described as follows:

$$p(D_j \mid \theta) = MAM_3(x_j, y_j \mid p, \mu, \sigma)$$

$$= \frac{p}{\sqrt{2\pi x_j^2}} e^{\frac{(y_j - \mu)^2}{2x_j^2}} + \frac{2(1-p)[1 - \Phi(\frac{\mu - \mu_0}{\sigma_0})]}{\sqrt{2\pi(x_j^2 + \sigma^2)}} e^{\frac{(y_j - \mu)^2}{2(x_j^2 + \sigma^2)}} \quad (2.2)$$

where

$$\mu_0 = \frac{\frac{\mu}{\sigma^2} + \frac{y_j}{x_j^2}}{\frac{1}{\sigma^2} + \frac{1}{x_j^2}} \text{ and } \sigma_0 = \frac{1}{\sqrt{\frac{1}{\sigma^2} + \frac{1}{x_j^2}}}.$$

Three quantities (p, μ, and σ) must be optimized in this model. p denotes the proportion of grains that are fully bleached before burial, μ denotes the logged characteristic dose, and σ is an adjustable parameter used to account for the dispersion in partially bleached grains (in unit per cent). Note that, conventionally, the logged characteristic dose in the MAM$_3$ is represented by γ, and in this study μ is alternatively used to make it correspond to that which is in the CAM. $\Phi(x)$ denotes the standard

Normal cumulative distribution function. The optimal parameters are typically estimated by direct numerical maximization of the joint-likelihood that is determined using paired observations $D=\{x_j, y_j \mid j=1,2,3,...,n)$. In contrast to the CAM, optimization of the MAM_3 is considered an ill-conditioned mathematical problem due to the strong nonlinearities in the calculation. Whether it is possible to converge it to global optimal estimates depends heavily on the choice of the initial parameters used. It is indispensable to try various initial parameter sets to obtain optimal estimates.

The maximum age model (MXAM)

The maximum age model (MXAM) is an analogue of the MAM and is firstly applied by Olley *et al.* (2006) in their estimation of the maximum dose population from D_e distributions. Galbraith and Roberts (2012) enumerated two scenarios where the MXAM is applicable: (1) Samples composed of fully bleached grains that have subsequently been mixed with younger, intrusive grains, and (2) samples composed of fully bleached grains, some proportion of which have been bleached after deposition.

The characteristic dose of the $MXAM_3$ can be estimated according to the MAM_3 by using a "mirror image" of the original D_e distribution as an input (Olley *et al.*, 2006; Galbraith and Roberts, 2012). Alternatively, by accounting for the fact that the lower truncation point of a truncated Normal distribution denotes the minimum characteristic dose in the MAM_3 while the upper truncation point of a truncated Normal distribution denotes the maximum characteristic dose in the $MXAM_3$, the likelihood for the $MXAM_3$ can be derived by a minor modification of **Eq. 2.2** and can be described as follows:

$$p(D_j \mid \theta) = MXAM_3(x_j, y_j \mid p, \mu, \sigma)$$

$$= \frac{p}{\sqrt{2\pi x_j^2}} e^{-\frac{(y_j-\mu)^2}{2x_j^2}} + \frac{2(1-p)\Phi(\frac{\mu-\mu_0}{\sigma_0})}{\sqrt{2\pi(x_j^2+\sigma^2)}} e^{-\frac{(y_j-\mu)^2}{2(x_j^2+\sigma^2)}} \quad (2.3)$$

where μ_0 and σ_0 are calculated in the same manner as the corresponding values in the MAM_3.

Unlogged versions of statistical age models

Larger D_e estimate typically have larger absolute standard errors; namely, larger D_e estimates will vary more (Arnold and Roberts, 2009). Consequently, the frequency distribution of D_e estimates will tend to be positively skewed. The Lognormal distribution provides a simple model to describe variables wherein their standard deviations increase in proportion to their means. It is for this reason that the aforementioned statistical age models use log-transformed D_e datasets as inputs. Datasets in log-scale are obtained by calculating natural logarithms of D_e estimates and transforming their absolute standard errors into RSEs. This is because the RSE of a D_e estimate approximates the standard error of a D_e estimate on

a natural logarithm scale (Galbraith and Roberts, 2012). As a result, these statistical age models are unsuitable for measured zero or negative D_e values that are consistent with zero or positive D_e values within a level of uncertainty. However, such types of D_e distribution are frequently encountered for either very young or modern samples. As a consequence, Arnold *et al.* (2009) used unlogged versions of the CAM and the MAM to analyze young/modern D_e datasets using a series of known-age fluvial samples and demonstrated that the unlogged versions of these models are capable of producing accurate burial dose estimates from their samples. To apply unlogged versions of these models, the user needs only use the original D_e values and their absolute standard errors as y_js and x_js, respectively, in **Eqs. 2.1–2.3**. In this case, the estimated μ denotes the characteristic dose (not the logged counterpart), while σ denotes a parameter in unit Gy (not in per cent).

3. ANALYZING GALBRAITH'S STATISTICAL AGE MODELS USING MCMC

MCMC sampling from statistical age models for individual samples

The statistical age models described in Section 2 are in fact equivalent dose models. This is because they estimate the characteristic D_e instead of the burial age (e.g., Guérin *et al.*, 2017). However, they can be easily transformed to obtain the posterior distribution of ages by taking the dose rate data into consideration during the simulation process:

$$\mu = a \times (\bar{d} + \varepsilon) \quad (3.1)$$

where a denotes the age, μ denotes the characteristic dose, $\bar{d}$ denotes the mean annual dose rate, and ε denotes a variable that follows Gaussian distribution with mean 0 and standard deviation σ_d, namely, $\varepsilon \sim N(0, \sigma_d^2)$. σ_d is calculated by aggregating many error sources due to measurements of uranium, thorium, potassium, water contents, *etc.* (Combès and Philippe, 2017). In this way, μ is represented mathematically using **Eq. 3.1**, and it is treated as an intermediate stochastic node during the simulation process. Consequently, the model can be re-parameterized using a instead of μ, namely, $\theta=[\mu, \sigma]$ (in the CAM) and $\theta=[p, \mu, \sigma]$ (in the MAM_3 or the $MXAM_3$) are re-parameterized as $\theta=[a, \sigma]$ and $\theta=[p, a, \sigma]$, respectively.

Each parameter is assigned a prior distribution that represents the prior knowledge of it before some evidence is considered. For p (*i.e.*, the proportion of fully bleached grains), a Uniform prior that lies between 0.001 and 0.999 was used. For a (*i.e.*, burial age) a typically employed non-informative prior is a Uniform distribution with lower and upper bounds set to equal a_{min} and a_{max}, which specifies a particular period of the burial history (e.g., Combès and Philippe, 2017). Parameter σ (*i.e.*, a

scale parameter inside the model) is an adjustable parameter that captures the unknown spread within the whole D_e distribution (for the CAM) or only the unknown spread in the D_e distribution for partially bleached grains (for the MAM₃). A significant amount of variation in OD exists between samples that either have different inherent luminescent properties, undergo different burial histories, and those that have been affected by various factors related to variation in microdosimetry and uncertainty from the measurement procedure. A summary of OD values from sedimentary samples (some are thought to have been fully bleached at the time of deposition) in published studies reveals that single-grain D_e distributions exhibit OD values of up to 30% (or higher), and the D_e distributions of multi-grain aliquots have a mean OD of 14% (Arnold and Roberts, 2009). OD values for several partially bleached young/modern fluvial samples calculated using an unlogged CAM range from 24% to 208% (Arnold *et al.*, 2009). In this study, the prior of σ was represented using a Uniform distribution over the interval (0, 5).

Galbraith and Roberts (2012) reiterated that an estimate of additional uncertainty (σ_b) should be added in the quadrature to the RSE of each D_e value before running the MAM (as well as the MXAM) (Galbraith *et al.*, 2005). σ_b represents the underlying spread in dose distribution typically present in well-bleached and undisturbed samples. An overestimate of σ_b will lead to an overestimate of the characteristic dose (and hence the burial age), and vice versa. Ideally, σ_b should be calculated from a well-bleached sample of the same mineral derived from the same source and, also ideally, of equivalent age (Galbraith and Roberts, 2012). In the absence of such information, σ_b values of 10% and 20% could be coarse approximations at multiple-grain and single-grain levels, respectively (e.g., Arnold and Roberts, 2009).

MCMC sampling from statistical age models for multiple samples with order constraints

In this section, an age-depth model to combine D_e datasets from multiple samples with stratigraphic constraints was considered by using the statistical age models described in Section 2. Although Bayesian age-depth models are expected to improve the precision of dates, such precision may be misleading if potential systematic uncertainties have not been included by the individual samples (Zeeden *et al.*, 2018). Compared to the application of statistical age models to individual samples using the standard approach, incorporating statistical age models into a Bayesian age-depth framework to refine the age-depth relationship between dates from multiple samples with order constraints requires that both random and systematic sources of errors are properly accounted for; otherwise, the precision of dates may increase decisively in unjustified ways (Combès and Philippe, 2017; Zeeden *et al.*, 2018).

This study adopted the method described by Combès and Philippe (2017) to build the relationship between the

characteristic dose μ_i, the annual dose rate $\bar{d}_i$, and the corresponding burial age a_i for the ith sample by taking into account both systematic and random sources of errors using the following Gaussian error model:

$$\mu_i = a_i \times (\bar{d}_i + \varepsilon_{di} + \varepsilon_{dc}) \tag{3.2}$$

where ε_{di} and ε_{dc} are the random measurement error for the ith sample and the systematic error shared by all samples from the sequence, respectively. The systematic error, which represents the error in the calibration of the measurement device (assuming that the measurements are made in the same luminescence laboratory), affects all estimates in the same way. The noise terms ε_{di} and ε_{dc} are assumed to follow Gaussian distributions, namely, $\varepsilon_{di} \sim N(0, \sigma_{di}^2)$, and $\varepsilon_{dc} \sim N(0, \sigma_{dc}^2)$, respectively. Therefore, the characteristic dose μ_i in the likelihoods is reparameterized using a combination of burial age a and the dose rate data (*i.e.*, the annual dose rate, associated random and systematic errors) according to **Eq. 3.2** and can be treated as an intermediate stochastic node.

For parameters p_i and σ_i of the ith sample (i=1, 2, 3, ..., N), the priors were set equal to p and σ, respectively. For parameter a_i, a Uniform distribution for which the lower and upper bounds were set equal to a_{mini} and a_{maxi} was used (the same as Section 3 – MCMC sampling from statistical age models for indi-vidual samples). In addition, a constraint on vector $[a_1, a_2, a_3, ..., a_N]$ such that $a_1 < a_2 < a_3 < ... < a_N$ (*i.e.*, wherein ages are sorted in ascending order) was imposed to allow stratigraphic information to be incorporated during the MCMC sampling process.

Stan (programming language) implementation for MCMC sampling

The objective being to obtain sampling distributions of parameters p(θ|D) based on the joint-likelihood p(D|θ) and priors $p(\theta)$ using MCMC sampling, a class of algorithms that utilize Markov chains to allow for the approximation of posterior distributions for parameters of interest by utilizing the prior distributions and likelihood as inputs (Annis *et al.*, 2017). This study adopted the recently developed software package Stan (Gelman *et al.*, 2015) for this purpose. Although there are dozens of built-in probability distributions that can be used to sample standard distributions (such as the Normal distribution, the Exponential distribution, *etc.*), users will occasionally require "non-standard" distributions (*i.e.*, the joint-likelihoods for the CAM, MAM₃, and MXAM₃ used in this study) that may not yet be defined. Annis *et al.* (2017) gave a detailed tutorial on the implementation of MCMC sampling using user-defined distributions within the Stan programming language. Once the user-supplied function is defined, it can be called upon as a built-in distribution. This section provides a simple description of the numeric programs used to conduct MCMC sampling from statistical age models applying D_e datasets from both individual and multiple samples using the Stan soft-

ware package. The relevant code and numeric scripts are freely downloadable from https://github.com/pengjunUCAS/mcmcSAM.

Individual samples

The input, likelihood, transformation, and distribution used to implement MCMC sampling for individual samples using the CAM and the MAM$_3$ are illustrated in **Fig. 1**. The MXAM$_3$ can be simulated in the same manner as the MAM$_3$ by changing the likelihood correspondingly. In **Fig. 1**, z_j and σ_{zj} denote the D$_e$ and associated absolute error for the jth aliquot (grain). These statements can easily be modified to accommodate unlogged versions of the CAM, MAM$_3$, and MXAM$_3$. In Stan, variable definitions, distribution specifications, and program statements are placed within code blocks (Annis *et al.*, 2017). Script "SAM0.stan" was employed to place all code blocks used to implement the sampling process. The R package "rstan" (an R interface for Stan) (Stan Development Team, 2018) was used to execute the script using a wrapper function mcmcSAM() available in the file "SAM.R". Once the sampling process is terminated, the resultant MCMC convergence diagnostics are automatically stored in the file "mcmcSAM.pdf", and the posterior samples are stored in the file "mcmcSAM.csv". Both files are available from current R working directory. Please refer to the file "SAM.R" for more detail on arguments and default settings used in function mcmcSAM().

Multiple samples with order constraints

Samples from the same sequence may substantially vary in D$_e$ distributions due to differences in provenance, depositional environment, microdosimetry, *etc.* There is no generally applicable statistical age model that can be used to analyze all D$_e$ distribution types. If we conceptualize three samples (S$_1$, S$_2$, and S$_3$) from the same strati-

graphic sequence with their sample depths in ascending order, and we assume that these samples exhibit significantly different D$_e$ distribution characteristics for which the most suitable models for these samples would be the MAM$_3$, CAM, and MXAM$_3$, respectively, then the statements used to conduct MCMC sampling on these samples applying a combination of different age models in log-scale can be summarized as shown in **Fig. 2**. z_{ij} and σ_{zij} denote the D$_e$ and the associated absolute standard error for the jth aliquot (grain) from the ith sample. y_{ij} and x_{ij} denote the logged D$_e$ and corresponding RSE for the jth aliquot (grain) from the ith sample. n_1, n_2, and n_3 denote the numbers of aliquots (grains) for the three samples. $p(D|\theta)$ is calculated as the product of joint likelihoods for samples S$_1$, S$_2$, and S$_3$ that are fitted using the MAM$_3$, CAM, and MXAM$_3$, respectively. Note that in this study only three samples are used for illustrative purposes, while in practice many samples fitted using various combinations of the three different age models can be simultaneously analyzed in a manner similar to that demonstrated in **Fig. 2**. This allows the programs to have a degree of flexibility when D$_e$ distributions from the same sequence need to be treated differently. Two scripts (*i.e.*, "SAMseq0.stan" and "SAMseq1.stan") were used to conduct the sampling process without and

CAM	MAM$_3$
for j in $1:n$	for j in $1:n$
$x_j = \sigma_{zj} / z_j$	$x_j = \sqrt{\sigma_{zj}^2 / z_j^2 + \sigma_b^2}$
$y_j = \log(z_j)$	$y_j = \log(z_j)$
end for	end for
$p(D|\theta) = \prod_{j=1}^{j=n} CAM(x_j, y_j | \mu, \sigma)$	$p(D|\theta) = \prod_{j=1}^{j=n} MAM_3(x_j, y_j | p, \mu, \sigma)$
$\mu = a \times d$	$\mu = a \times d$
$d \sim Normal(\bar{d}, \sigma_d)$	$d \sim Normal(\bar{d}, \sigma_d)$
$a \sim Uniform(a_{min}, a_{max})$	$p \sim Uniform(0.001, 0.999)$
$\sigma \sim Uniform(0, 5)$	$a \sim Uniform(a_{min}, a_{max})$
	$\sigma \sim Uniform(0, 5)$

Fig. 1. *Statements used to conduct MCMC sampling according to the CAM and the MAM$_3$ (log-scale) using D$_e$ sets from individual samples. The MXAM$_3$ can be applied in a similar manner to the MAM$_3$ by changing the likelihood function.*

Apply MAM$_3$, CAM, and MXAM$_3$ to three samples with order constraints
for i in $1:3$
for j in $1:n_i$
$x_{ij} = \sqrt{\sigma_{zij}^2 / z_{ij}^2 + \sigma_{bi}^2}$
$y_{ij} = \log(z_{ij})$
end for
end for
$p(D|\theta) = \prod_{j=1}^{j=n_1} MAM_3(x_{1j}, y_{1j} | p_1, \mu_1, \sigma_1) \times$
$\prod_{j=1}^{j=n_2} CAM(x_{2j}, y_{2j} | \mu_2, \sigma_2) \times$
$\prod_{j=1}^{j=n_3} MXAM_3(x_{3j}, y_{3j} | p_3, \mu_3, \sigma_3)$
for i in $1:3$
$\mu_i = a_i \times d_i$
$d_i \sim Normal(\bar{d}_i, \sigma_{di}^2 + \sigma_{dc}^2)$
$p_i \sim Uniform(0.001, 0.999)$
$a_i \sim Uniform(a_{min i}, a_{max i}) \in [a_1 < a_2 < a_3]$
$\sigma_i \sim Uniform(0, 5)$
end for

Fig. 2. *Statements used for conducting MCMC simulation simultaneously according to different age models (log-scale) using D$_e$ sets from three samples whose ages were constrained such that $a_1 < a_2 < a_3$. The three samples were analyzed using the MAM$_3$, CAM, and MXAM$_3$, respectively. The protocol can be easily adapted for MCMC sampling without order constrains for burial ages.*

with order constraints, respectively. A wrapper function `mcmcSAMseq()` that is available in file "SAMseq.R" was used to call these scripts. The resultant MCMC convergence diagnostics are automatically stored in file "mcmcSAMseq.pdf", and the posterior samples are stored in file "mcmcSAMseq.csv".

4. APPLICATIONS AND RESULTS

Measured D_e sets

First, the validity of the MCMC sampling protocol (as demonstrated in **Fig. 1**) used to analyze D_e sets from individual samples was checked using measured multi-grain D_e distributions (**Fig. 3**) of four aeolian samples (*i.e.*, GL2-1 to GL2-4) from Gulang County at the southern margin of the Tengger Desert, China (Peng *et al.*, 2016b). The D_e distributions were obtained using the single aliquot regenerative-dose (SAR) protocol (Galbraith *et al.*, 1999; Murray and Wintle, 2000). When

applying the MAM$_3$ and MXAM$_3$, an additional uncertainty (*i.e.*, σ_b) of 10% was added in the quadrature to the RSEs of the measured D_e values to account for unrecognized measurement errors. In this study (unless stated otherwise), three parallel Markov chains (the number of iterations for each chain was set equal to 10000) were simulated during the MCMC sampling process for the individual samples, and four parallel Markov chains (the number of iterations for each chain was set equal to 6000) were simulated during the MCMC sampling process for the multiple samples. The lower and upper bounds for the priors of ages were set equal to $a_{min}(a_{mini})$=1 ka and $a_{max}(a_{maxi})$=100 ka, respectively. The characteristic dose (*i.e.*, μ) and age (*i.e.*, a) of the CAM, MAM$_3$, and MXAM$_3$ obtained using MLE and MCMC were then compared and the results summarized in **Table 1**. The estimates obtained between the two different methods were consistent within errors, suggesting that estimates derived from MCMC can be used as an informative comparison for those estimated by MLE.
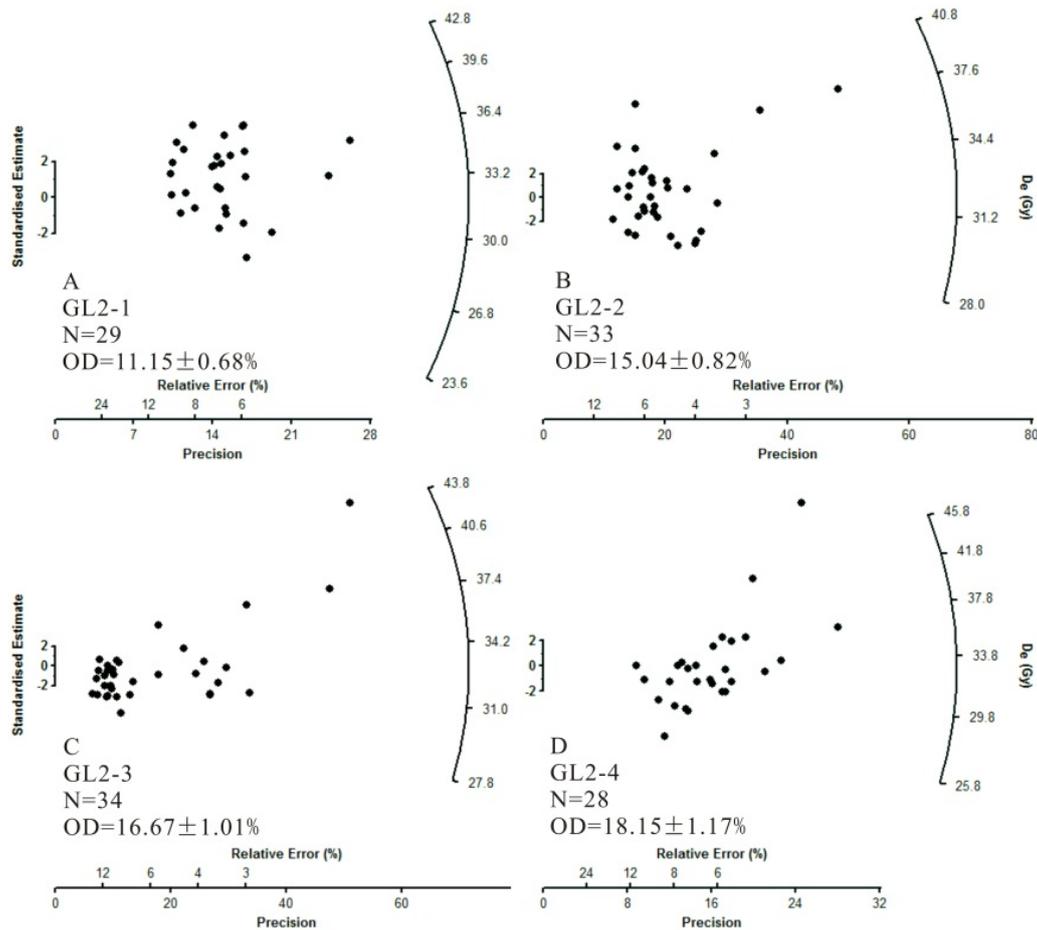
Fig. 3. *Measured multi-grain D_e distributions for four aeolian samples from Gulang County at the southern margin of the Tengger Desert, China (Peng et al., 2016b). N denotes the number of measured aliquots. Note that in Peng et al. (2016b) an additional uncertainty of σ_b=5% was added in the quadrature to the RSEs of the measured D_e values to account for sources of errors that were not considered in the measurements. However, no additional uncertainty was added to the D_e distributions shown here.*

These four samples were also analyzed simultaneously using the MCMC sampling protocol (as demonstrated in **Fig. 2**) for multiple samples. D_e values of these samples fell within reasonably narrow ranges and displayed homogeneous distributions (**Fig. 3**). Accordingly, the CAM was applied to determine the burial dose. Samples GL2-1, GL2-2, GL2-3, and GL2-4 were collected from the same section in ascending order depths. Consequently, their burial ages were also expected to be in ascending order. The MCMC sampling protocol was applied to these samples with order constraints (results shown in **Table 2**). The systematic error representing the error in the calibration of the dose rate measurement device shared by all samples was assumed to be σ_{dc}=0.1. The posterior distributions of burial age are shown in **Fig. 4**. Results obtained without order constraints were also provided for comparison. These results suggested that burial ages would be internally consistent with the stratigraphy if order constraints were imposed during the MCMC sampling process. Moreover, the posterior standard deviations of burial ages would be significantly reduced when using stratigraphic constraints, implying that the precision of burial ages improved. Finally, a comparison between **Table 1** and **Table 2** for CAM results demonstrated that burial dose and ages (and their posterior standard deviations) obtained by MCMC sampling using both individual samples and multiple samples simultaneously without order constraints were indistinguishable from each other.

Simulated D_e sets

This section shows simulation results of a series of D_e distributions with known burial ages to further verify the performance of the MCMC sampling protocols. The first three samples (#1–#3) were assumed to be partially-bleached, the next three samples (#4–#6) fully-bleached, and the last three samples (#7–#9) contain fully-bleached grains that have been subsequently mixed with younger, intrusive grains. These samples were assumed to be collected from the same sedimentary sequence but at different depths. Their "true" burial doses ranged from 18 to 42 Gy with a step of 3 Gy, and their "true" burial ages ranged from 7 to 11 ka with a step of 0.5 ka. The "true" average dose rate can be obtained by dividing the "true" burial dose by the "true" burial age of a sample. The random error arising from dose rate measurement was assumed to be 7%, in order to simulate a realistic annual dose rate. Namely, the "true" dose rates are noised using a Normal distribution with a mean equivalent to the "true" average dose rate and a relative standard deviation (RSD) equal to 7%. The generated random dose rates were used as the measured dose rates when determining burial ages using MLE and MCMC. The systematic error representing the error in the calibration of the dose rate measurement device shared by all samples was assumed to be zero (*i.e.*, σ_{dc}=0).

Lognormal distributions were used to model the natural dispersion in dose distributions arising from dose rate variation with an RSD of 5% (*i.e.*, σ_d=0.05). Measured D_e

Table 1. *Comparisons of burial dose and age estimated from MLE and MCMC for the various statistical age models using D_e sets from four measured aeolian samples. Quantities estimated using the MCMC sampling protocol shown in* Fig. 1 *are marked in bold. Quantities estimated using the MLE are inside parentheses.*

Sample	Dose rate (Gy/ka)	CAM		MAM₃		MXAM₃	
		μ (Gy)	a (ka)	μ (Gy)	a (ka)	μ (Gy)	a (ka)
GL2-1	3.26 ± 0.25	**34.55 ± 0.91**	**10.73 ± 0.89**	**33.96 ± 1.17**	**10.55 ± 0.89**	**35.36 ± 1.41**	**10.96 ± 0.97**
		(34.51 ± 0.84)	(10.58 ± 0.85)	(32.79 ± 3.64)	(10.05 ± 1.36)	(35.98 ± 3.51)	(11.04 ± 1.37)
GL2-2	2.84 ± 0.21	**32.70 ± 0.98**	**11.64 ± 0.95**	**30.40 ± 1.41**	**10.83 ± 0.98**	**35.53 ± 2.12**	**12.66 ± 1.26**
		(32.65 ± 0.91)	(11.50 ± 0.91)	(30.33 ± 1.67)	(10.68 ± 0.98)	(37.12 ± 2.31)	(13.7 ± 1.26)
GL2-3	3.23 ± 0.23	**30.57 ± 1.05**	**9.56 ± 0.77**	**27.92 ± 1.55**	**8.74 ± 0.80**	**35.19 ± 2.03**	**11.00 ± 1.01**
		(30.55 ± 0.98)	(9.46 ± 0.74)	(26.03 ± 3.56)	(8.06 ± 1.24)	(36.61 ± 2.41)	(11.33 ± 1.10)
GL2-4	3.40 ± 0.24	**32.23 ± 1.27**	**9.58 ± 0.79**	**30.09 ± 1.55**	**8.94 ± 0.80**	**37.77 ± 2.61**	**11.22 ± 1.12**
		(32.17 ± 1.18)	(9.46 ± 0.75)	(30.75 ± 1.04)	(9.04 ± 0.71)	(39.16 ± 2.46)	(11.52 ± 1.09)

Table 2. *A summary of burial dose and age estimated from the CAM for D_e sets from four measured aeolian samples using the MCMC sampling protocol shown in* Fig. 2 *without and with order constraints. The systematic error for the dose rate measurement shared by all samples was set to σ_{dc}=0.1.*

Sample	Dose rate (Gy/ka)	MCMC (without constraints)		MCMC (with constraints)	
		μ (Gy)	a (ka)	μ (Gy)	a (ka)
GL2-1	3.26 ± 0.25	34.55 ± 0.91	10.75 ± 0.96	34.23 ± 0.86	9.72 ± 0.52
GL2-2	2.84 ± 0.21	32.71 ± 1.00	11.68 ± 1.06	32.19 ± 0.92	10.28 ± 0.49
GL2-3	3.23 ± 0.23	30.58 ± 1.04	9.59 ± 0.85	31.03 ± 1.00	10.59 ± 0.52
GL2-4	3.40 ± 0.24	32.24 ± 1.29	9.60 ± 0.85	33.13 ± 1.26	11.13 ± 0.67
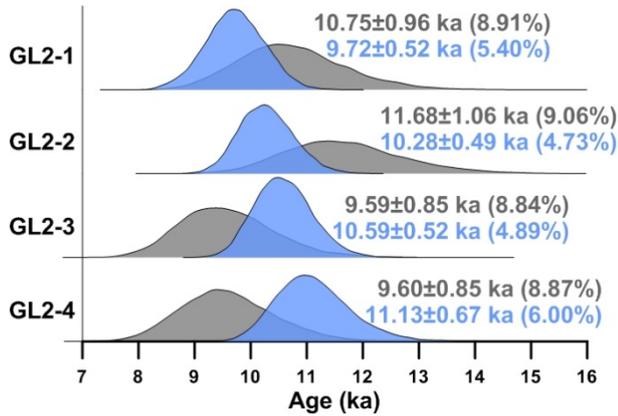
Fig. 4. *Posterior distributions of burial ages obtained from the CAM for the four aeolian samples taken from the same sedimentary section obtained using the MCMC sampling protocol, as shown in Fig. 2. The blue-coloured density plots denote results obtained by constraining the burial ages in ascending order. The grey-coloured density plots denote results obtained by imposing no constraints on the order of burial ages. The value inside parentheses denotes the calculated RSD of burial age.*

Fig. 5. *Distribution of OSL sensitivity for the 127 grains of a sample taken from Lake Mungo, Australia. The red line denotes the probability density curve obtained by fitting the OSL sensitivity data using a Gamma distribution. The fitted Gamma distribution has a shape parameter of α=0.565 and a rate parameter of β=0.191. Q[x%] denotes the x% sample quantile of OSL sensitivity.*

distributions were simulated according to the method described by Li *et al.* (2017), which involves simulating OSL intensities of the natural dose (L_n), a series of regenerative doses (L_x), and their corresponding test doses (T_n and T_x). Measured OSL sensitivity data taken from an empirical sample was used as the basis of the simulation. The OSL sensitivity (e.g., counts per unit of dose) of the different grains were randomly generated from a Gamma distribution fitted to the single-grain T_n datum from a sample taken from Lake Mungo, Australia (**Fig. 5**). Noise was added to each of the simulated OSL signal intensities based on the uncertainty arising from counting statistics and instrumental irreproducibility. It was assumed that both photon counts and dark counts were over-dispersed compared to a Poisson distribution (e.g., Li, 2007; Adamiec *et al.*, 2012) and follow Negative Binomial distributions (e.g., Bluszcz *et al.*, 2015). The variance to mean ratios for the photon counts and dark counts were set equal to 1.38 and 1.85, respectively. The dark count was set equal to 10 cts/0.2s. The instrumental reproducibility was set as 2%. For the natural signals (L_n), an extra noise from "intrinsic OD" (e.g., σ_m=10%), which represents unrecognized measurement errors (e.g., OD observed from a dose recovery experiment) (e.g., Thomsen *et al.*, 2005; Jacobs *et al.*, 2006; Guérin *et al.*, 2017), was also added. D_e estimation was conducted using the R package "numOSL" (Peng *et al.*, 2013; Peng and Li, 2017). For each sample, 100 D_e values (and associated errors) were simulated. The nine simulated D_e distributions are shown in **Fig. 6**.

The simulated D_e sets were analyzed individually using MLE, analyzed simultaneously using MCMC without order constraints, and analyzed simultaneously using MCMC with order constraints. The first, next, and last

three samples were analyzed using the MAM$_3$, CAM, and MXAM$_3$, respectively. To facilitate the model in the analysis of the MAM$_3$ and the MXAM$_3$, a σ_b should be determined beforehand. The calculated OD values for samples #4–#6 can be regarded as appropriate σ_b values, given that these samples are "fully-bleached" simulated samples whose errors, arising from dose rate variation (*i.e.*, σ_d) and unrecognized measurement uncertainty (*i.e.*, σ_m), were the same as samples #1–#3 and samples #7–#9. The estimated OD values were 10.09±0.31%, 10.48±0.33%, and 11.89±0.38%, respectively, for samples #4–#6. These values are broadly consistent with the expected value of $\sqrt{\sigma_d^2 + \sigma_m^2} = \sqrt{(5\%)^2 + (10\%)^2} = 11.18\%$, confirming the validity of the simulation. Accordingly, a σ_b value of 11% was added to the RSEs of the D_e values when analyzing the MAM$_3$ and the MXAM$_3$. Analyzed results for each sample are summarized in **Table 3**. The posterior distributions of burial ages are shown in **Fig. 7** for results obtained by MCMC sampling with and without order constrains. Similar to Section 4 (Measured D_e sets), the burial doses and ages (and associated errors) obtained by MLE using individual samples and by MCMC using multiple samples simultaneously without order constraints were broadly consistent between one another. When order constraints were imposed during the MCMC sampling process, the precision of burial age improved for all samples, and the accuracy of burial age also increased for most samples (except for samples #4 and #9) (**Fig. 7**). The relative error (RE) of

Table 3. *A summary of burial dose and age estimated from various statistical age models for nine simulated D_e sets taken from the same sedimentary sequence. Results obtained using MLE and results obtained using the MCMC sampling protocol as shown in Fig. 2 without and with order constraints are presented. The systematic error on the dose rate measurement shared by all simulated samples was assumed to be $\sigma_{dc}=0$.*

Sample	Model	Age (ka)	Burial dose (Gy)	MLE		MCMC (without constraints)		MCMC (with constraints)	
				μ (Gy)	a (ka)	μ (Gy)	a (ka)	μ (Gy)	a (ka)
#1		7	18	18.35 ± 1.17	7.95 ± 0.75	17.85 ± 1.17	7.82 ± 0.76	16.58 ± 1.02	6.80 ± 0.44
#2	MAM₃	7.5	21	21.16 ± 0.63	6.95 ± 0.53	20.96 ± 0.79	6.96 ± 0.57	21.17 ± 0.65	7.23 ± 0.40
#3		8	24	24.47 ± 1.22	8.41 ± 0.72	23.64 ± 1.57	8.22 ± 0.81	23.25 ± 1.43	7.84 ± 0.42
#4		8.5	27	26.91 ± 0.33	8.24 ± 0.59	26.92 ± 0.34	8.32 ± 0.61	26.93 ± 0.33	8.31 ± 0.39
#5	CAM	9	30	30.40 ± 0.38	8.50 ± 0.60	30.41 ± 0.39	8.59 ± 0.63	30.44 ± 0.38	8.83 ± 0.42
#6		9.5	33	32.79 ± 0.46	10.59 ± 0.76	32.81 ± 0.47	10.71 ± 0.79	32.66 ± 0.45	9.62 ± 0.42
#7		10	36	35.77 ± 1.61	9.13 ± 0.76	36.43 ± 1.86	9.39 ± 0.84	37.18 ± 1.81	9.95 ± 0.45
#8	MXAM₃	10.5	39	39.26 ± 1.37	9.32 ± 0.73	39.70 ± 1.69	9.53 ± 0.79	40.62 ± 1.87	10.39 ± 0.54
#9		11	42	43.65 ± 2.55	10.63 ± 0.97	44.61 ± 2.52	10.96 ± 1.00	45.43 ± 2.53	11.48 ± 0.85
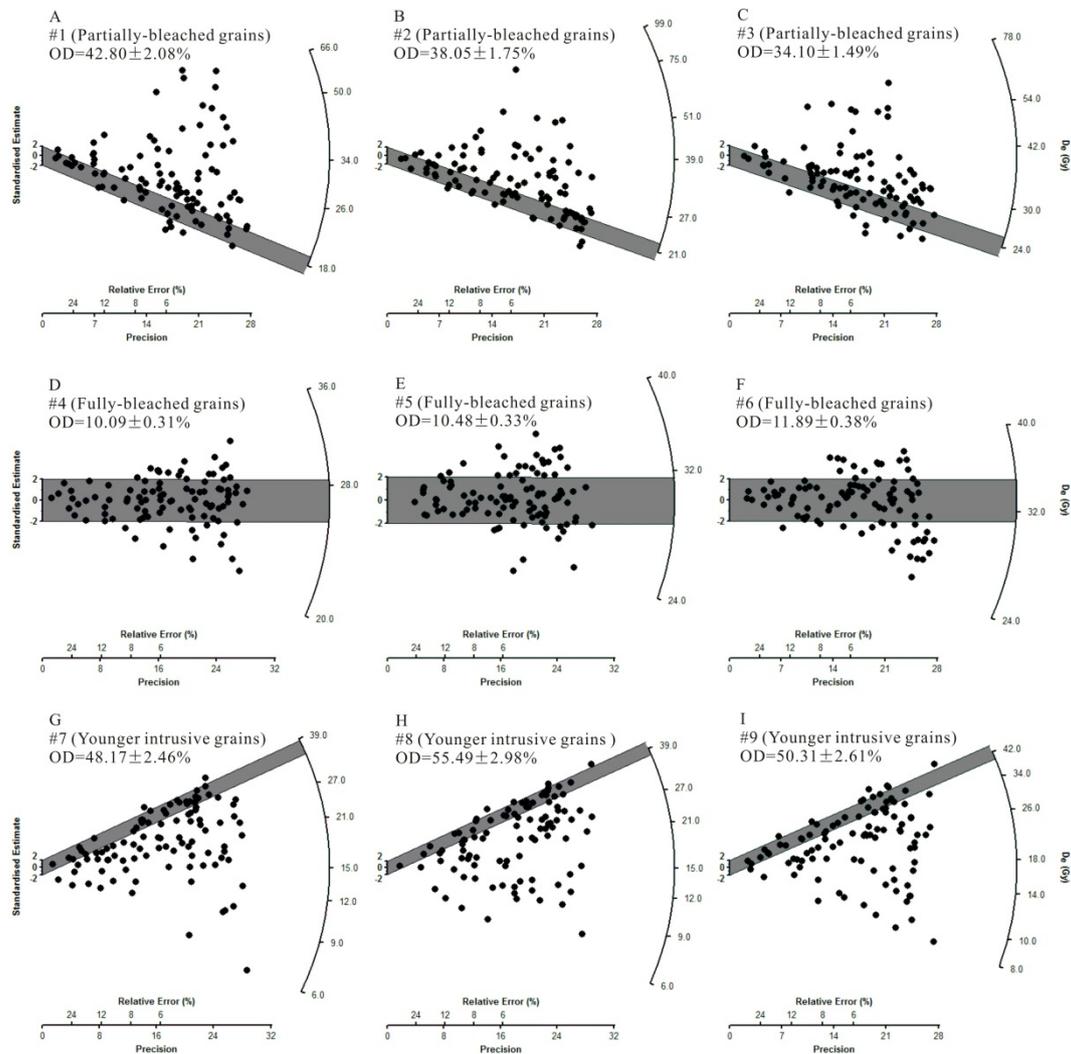
Fig. 6. *Simulated D_e distributions of nine samples from a sedimentary sequence. Each D_e set contains 100 D_e values (and associated standard errors). The shaded bars in each plot indicate the burial dose within 2σ error. The first three samples (#1–#3) are partially-bleached, the next three samples (#4–#6) are fully-bleached, and the last three samples (#7–#9) contain fully-bleached grains that have been subsequently mixed with younger, intrusive grains. The nine samples are assumed to be collected from the same section but at different depths, and their estimated burial dose and age are summarized in Fig. 7 and Table 3.*
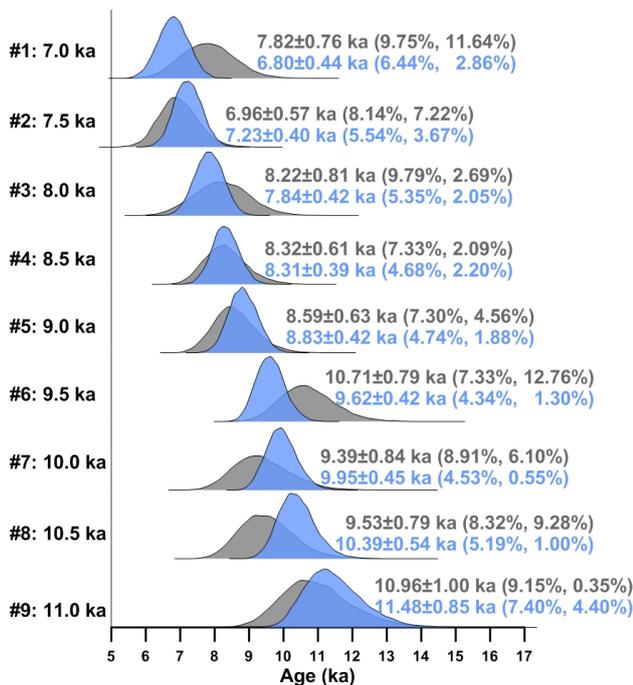
Fig. 7. *Posterior distributions of burial ages of the nine simulated samples from a sedimentary section obtained using the MCMC sampling protocol, as shown in* Fig. 2. *Samples #1–#3, #4–#6, and #7–#9 were analyzed using the MAM₃, CAM, and MXAM₃, respectively. The blue-coloured density plots are results obtained by constraining burial ages in ascending order. The grey-coloured density plots are results obtained by imposing no constraints on the order of burial ages. The first and second values inside parentheses denote the calculated RSD and RE of the burial age, respectively.*

burial age of sample #4 increased slightly from 2.09% to 2.20%, and that of sample #9 increased substantially from 0.35% to 4.4%, if order constraints were imposed.

MCMC convergence diagnostics

Although a properly constructed MCMC protocol enables one to draw samples successively from a convergent Markov chain, deciding the point to terminate sampling is sometimes difficult as well as the point to confidently conclude when the algorithm has converged to the desired stationary distribution. One simple and direct way to inspect the convergence of a chain is to check the degree of mixing of the chain by using a trace plot (**Fig. 8A**). Values that get "stuck" within certain space intervals of variables indicate poor mixing. In contrast, a chain that shows good mixing properties will move freely along the feasible space of variables. Another simple method for assessing the convergence of a chain is to monitor the autocorrelations of generated variables (**Fig. 8C**). Varia-

bles generated using MCMC suffer from autocorrelations to some extent. To decrease the autocorrelations of a variable, simulated samplers are routinely trimmed using a "thinning" protocol after the "burn-in" ("warm-up") procedure. If the autocorrelations of variables remain extremely high after applying a large "thinning" value, it may imply that the speed of the convergence is slow (*i.e.,* poor mixing) and more samples are therefore needed so that a meaningful inference can be drawn. In this study (unless stated otherwise), the number of "warm-up" iterations per chain was set equal to 2000, and the number of "thinning" iterations per chain was set equal to 1.

The Gelman-Rubin convergence diagnostic (Gelman and Rubin, 1992) is one of the most widely used methods for monitoring the convergence of MCMC outputs. Several parallel chains are simulated using various initial states when applying this method, and a shrink factor is used as a measurement of the difference between within-chain and between-chain variance (which is similar to the analysis of variance). A shrink factor value equal to or less than 1 indicates adequate convergence, while a factor value significantly above 1 indicates a lack of convergence. Results from Gelman-Rubin convergence diagnostics for posterior samples of CAM burial ages from sample GL2-1 is shown in **Fig. 8D**. It can be clearly observed that the simulations stabilized and were very close to 1 after approximately 3000 iterations. A summary of Gelman-Rubin convergence diagnostics for sequences of measured and simulated samples obtained using the MCMC sampling protocol of **Fig. 2** with order constraints is presented in **Table 4**.

5. DISCUSSION

This study adopted an MCMC method to obtain sampling distributions on parameters of interest in statistical age models, including the MAM₃, CAM, and MXAM₃, using D_e distributions from individual samples and multiple samples with order constraints. It demonstrated that the estimates obtained by MLE and MCMC from measured samples were consistent within errors for various age models (**Table 1**). In addition, the burial doses and ages obtained by MLE using individual samples and by MCMC using multiple samples simultaneously without order constraints were also broadly consistent between one another (**Tables 1–3**). These results are encouraging, indicating that the MCMC method potentially provides an alternative perspective for which to analyze statistical age models, and the reliability of parameters of interest can be assessed by comparing the results between the two independent methods (Peng *et al.*, 2016a).
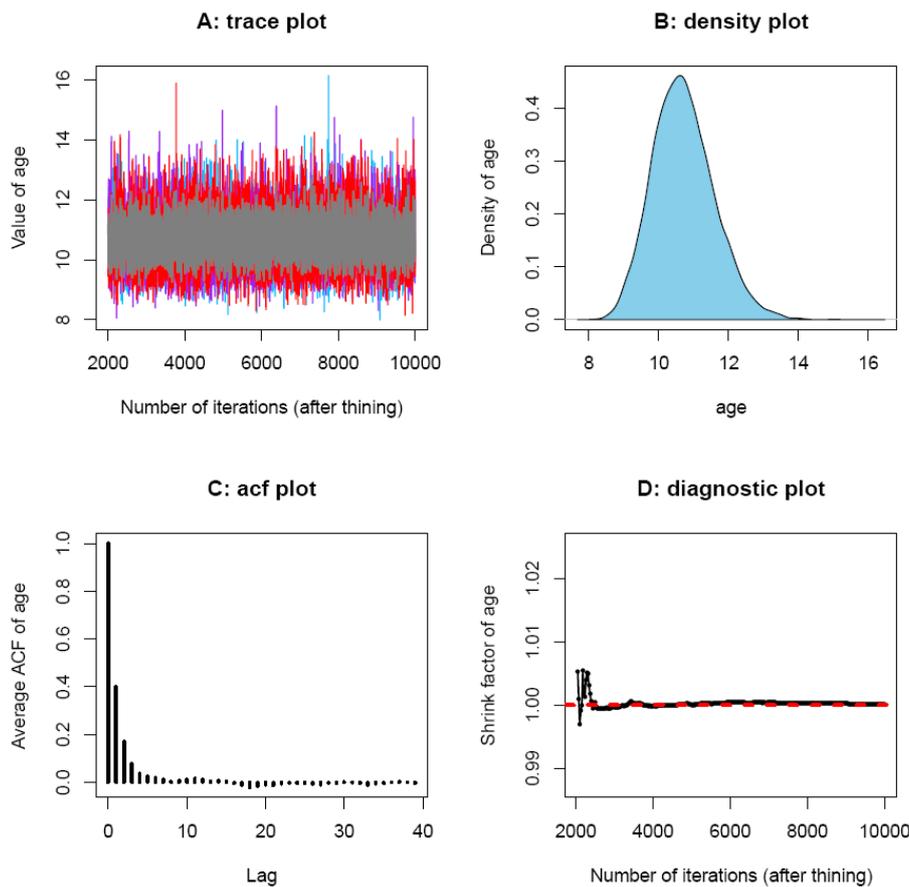
Fig. 8. *Convergence diagnostics for posterior samples of the CAM burial age from sample GL2-1. The plots were automatically generated using function* `mcmcSAM()`. *(A) denotes the trace plot; (B) denotes the density plot; (C) denotes the autocorrelations plot, and (D) denotes the Gelman-Rubin convergence diagnostic plot generated using three parallel Markov chains.*

Table 4. *A summary of Gelman-Rubin convergence diagnostics for measured and simulated samples obtained using the MCMC sampling protocol, as shown in Fig. 2 with order constraints. n_eff is the effective sample size, while Rhat (i.e., the shrink factor) is a statistic measure of the ratio of the average variance of samples within each chain to the variance of the pooled samples across chains. If all chains are at equilibrium, the Rhat will be 1. If these chains have not converged to a common distribution, the Rhat statistic will be greater than 1.*

Sample	μ		a		Sample	μ		a	
	n_eff	Rhat	n_eff	Rhat		n_eff	Rhat	n_eff	Rhat
GL2-1	16000	0.9998	5541.74	1.0002	#1	13865.37	1.0001	9987.93	1.0000
GL2-2	16000	0.9999	8220.18	1.0003	#2	16000	1.0001	11473.55	1.0000
GL2-3	16000	0.9997	7595.60	1.0002	#3	9405.95	1.0004	12466.84	1.0003
GL2-4	16000	0.9998	8864.01	1.0003	#4	16000	1.0003	13161.07	0.9999
					#5	16000	0.9999	12830.44	0.9999
					#6	16000	0.9999	14048.98	0.9998
					#7	16000	0.9999	13073.69	0.9998
					#8	7780.82	1.0003	12665.56	1.0001
					#9	16000	0.9999	16000	0.9999

For all MCMC simulations, initial states did not need to be specified and were generated automatically using random number generators, suggesting that the MCMC sampling method employed in this study was not sensitive to the choice of the initial state, and that the algorithm itself was free from the influence of local modes and was able to converge to the desired distribution with a finite number of iterations (**Fig. 8** and **Table 4**). This has a clear advantage over MLE, whereby the latter may result in very different estimates when various starting values are tried, especially for complex age models such as the MAM_3, and the $MXAM_3$ (Peng *et al.*, 2013). How-

ever, a major drawback of the MCMC method compared to MLE is that a large number of iterations (say, in the tens of thousands) are required in order to explore the entire range of the target distribution, and the process will increasingly be time-consuming given the increase in the number of observations or the complexity of the model under consideration.

When applying the MCMC sampling method to multiple samples with stratigraphic constraints using statistical age models, samples are not analyzed on a standalone basis; rather, they are combined simultaneously in a single model implementation so that between-sample age-depth relationships can also be included. This shows that the burial age of measured samples was internally consistent with the stratigraphy, and the posterior standard deviations of burial ages significantly decreased (**Fig. 4**). Using simulated D_e sets with known burial ages demonstrate that the MCMC sampling method can not only increase the precision but also may improve the accuracy of age estimates (**Fig. 7**). These results illustrate the benefit of including stratigraphic constraints when available during the analysis of statistical age models using D_e sets from multiple samples (e.g., Cunningham and Wallinga, 2012; Combès and Philippe, 2017). However, it should be noted that incorporating age-depth relationship between samples into the Bayesian model may also have the risk of decreasing the accuracy of age estimates for certain samples within the sequence, as observed in **Fig. 7**.

The sequence of D_e sets simulated in Section 4 (Simulated D_e sets) was determined to be free of systematic error arising from dose rate measurements. However, the systematic error term that induces dependences between observations and the associated age-depth model should always be taken into account when analyzing measured samples (e.g., Rhodes *et al.*, 2003; Combès and Philippe, 2017; Zeeden *et al.*, 2018), although in most cases this error term remains unresolved and subject to guesswork. For example, the systematic error on the dose rate measurement shared by all samples was set to $\sigma_{dc}=0.1$ for measured samples in this study. The quality of estimates for burial ages may potentially be improved by introducing external (independent) dates, allowing for the correction of systematic components of the error (e.g., Rhodes *et al.*, 2003). This function may be made available in future releases of relevant software programs.

6. CONCLUSIONS

This study employed an MCMC sampling method for the statistical analysis of D_e datasets, using statistical age models, including the CAM, MAM$_3$, and MXAM$_3$. The method was tested using measured, and simulated D_e sets, and consistent results in agreement with expected values were obtained. The development of numerical programs provides a statistical toolbox for the calculation of OSL ages using MCMC sampling for both individual samples and multiple samples with (or without) order constraints

by taking into account systematic and individual errors. The MCMC method employed in this study can be used in addition to other Bayesian models in tackling luminescence-related chronological data and can be flexibly adapted to compute OSL ages of both well- and poorly-bleached sedimentary samples.

REFERENCES

Adamiec G, Heer AJ and Bluszcz A, 2012. Statistics of count numbers from a photomultiplier tube and its implications for error estimation. *Radiation Measurements* 47(9): 746–751, DOI 10.1016/j.radmeas.2011.12.009.

Annis J, Miller BJ and Palmeri TJ, 2017. Bayesian inference with Stan: A tutorial on adding custom distributions. *Behaviour research methods* 49(3): 863–886, DOI 10.3758/s13428-016-0746-9.

Arnold LJ, Bailey RM and Tucker GE, 2007. Statistical treatment of fluvial dose distributions from southern Colorado arroyo deposits. *Quaternary Geochronology* 2(1–4): 162–167, DOI 10.1016/j.quageo.2006.05.003.

Arnold LJ and Roberts RG, 2009. Stochastic modelling of multi-grain equivalent dose (De) distributions: Implications for OSL dating of sediment mixtures. *Quaternary Geochronology* 4(3): 204–230, DOI 10.1016/j.quageo.2008.12.001.

Arnold LJ, Roberts RG, Galbraith RF and DeLong SB, 2009. A revised burial dose estimation procedure for optical dating of young and modern-age sediments. *Quaternary Geochronology* 4(4): 306–325, DOI 10.1016/j.quageo.2009.02.017.

Bayliss A and Ramsey CB, 2004. *Pragmatic Bayesians: a decade of integrating radiocarbon dates into chronological models*. Lecture Notes in Statistics, Springer, London. DOI 10.1007/978-1-4471-0231-1_2.

Bluszcz A, Adamiec G and Heer AJ, 2015. Estimation of equivalent dose and its uncertainty in the OSL SAR protocol when count numbers do not follow a Poisson distribution. *Radiation Measurements* 81: 46–54, DOI 10.1016/j.radmeas.2015.01.004.

Brill D, Jankaew K and Brückner H, 2015. Holocene evolution of Phra Thong's beach-ridge plain (Thailand) — Chronology, processes and driving factors. *Geomorphology* 245: 117–134, DOI 10.1016/j.geomorph.2015.05.035.

Christophe C, Philippe A, Guérin G, Mercie N and Guibert P, 2018. Bayesian approach to OSL dating of poorly bleached sediment samples: Mixture Distribution Models for Dose (MD^2). *Radiation Measurements* 108: 59–73, DOI 10.1016/j.radmeas.2017.10.007.

Combès B and Philippe A, 2017. Bayesian analysis of individual and systematic multiplicative errors for estimating ages with stratigraphic constraints in optically stimulated luminescence dating. *Quaternary Geochronology* 39: 24–34, DOI 10.1016/j.quageo.2017.02.003.

Combès B, Philippe A, Lanos P, Mercier N, Tribolo C, Guerin G, Guibert P and Lahaye C, 2015. A Bayesian central equivalent dose model for optically stimulated luminescence dating. *Quaternary Geochronology* 28: 62–70, DOI 10.1016/j.quageo.2015.04.001.

Cunningham AC, Evans M and Knight J, 2015b. Quantifying bleaching for zero-age fluvial sediment: A Bayesian approach. *Radiation Measurements* 81: 55–61, DOI 10.1016/j.radmeas.2015.04.007.

Cunningham AC and Wallinga J, 2012. Realizing the potential of fluvial archives using robust OSL chronologies. *Quaternary Geochronology* 12: 98–106, DOI 10.1016/j.quageo.2012.05.007.

Cunningham AC, Wallinga J, Hobo N, Versendaal AJ, Makaske B and Middelkoop H, 2015a. Re-evaluating luminescence burial doses and bleaching of fluvial deposits using Bayesian computational statistics. *Earth Surface Dynamics* 3(1): 55–65, DOI 10.5194/esurf-3-55-2015.

Galbraith RF and Green PF, 1990. Estimating the component ages in a finite mixture. *Nuclear Tracks and Radiation Measurements* 17(3): 197–206, DOI 10.1016/1359-0189(90)90035-V.

Galbraith RF and Roberts RG, 2012. Statistical aspects of equivalent dose and error calculation and display in OSL dating: An overview and some recommendations. *Quaternary Geochronology* 11: 1–27, DOI 10.1016/j.quageo.2012.04.020.

Galbraith RF, Roberts RG, Laslett GM, Yoshida H and Olley JM, 1999. Optical dating of single and multiple grains of quartz from Jinmium rock shelter, northern Australia: Part I, experimental design and statistical models. *Archaeometry* 41(2): 339–364, DOI 10.1111/j.1475-4754.1999.tb00987.x.

Galbraith RF, Roberts RG and Yoshida H, 2005. Error variation in OSL palaeodose estimates from single aliquots of quartz: a factorial experiment. *Radiation Measurements* 39(3): 289–307, DOI 10.1016/j.radmeas.2004.03.023.

Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A and Rubin DB, 2014. *Bayesian data analysis*. Boca Raton, FL, CRC press.

Gelman A, Lee D and Guo JQ, 2015. Stan: A probabilistic programming language for Bayesian inference and optimization. *Journal of Educational and Behavioral Statistics* 40(5): 530–543, DOI 10.3102/1076998615606113.

Gelman A and Rubin DB, 1992. Inference from iterative simulation using multiple sequences. *Statistical Science* 7(4): 457–472, DOI 10.1214/ss/1177011136.

Guérin G, Christophe C, Philippe A, Murray AS, Thomsen KJ, Tribolo C, Urbanova P, Jain M, Guibert P, Mercier N, Kreutzer S and Lahaye C, 2017. Absorbed dose, equivalent dose, measured dose rates and implications for OSL age estimates: Introducing the Average Dose Model. *Quaternary Geochronology* 41: 163–173, DOI 10.1016/j.quageo.2017.04.002.

Guérin G, Combès B, Lahaye C, Thomsen KJ, Tribolo C, Urbanova P, Guibert P, Mercier N and Vallada H, 2015. Testing the accuracy of a Bayesian central-dose model for single-grain OSL, using known-age samples. *Radiation Measurements* 81: 62–70, DOI 10.1016/j.radmeas.2015.04.002.

Huntriss A, 2008. *A Bayesian analysis of luminescence dating*. Doctoral dissertation, Durham University.

Jacobs Z, Duller GAT and Wintle AG, 2006. Interpretation of single grain De distributions and calculation of De. *Radiation Measurements* 41(3): 264–277, DOI 10.1016/j.radmeas.2005.07.027.

Kunz A, Pflanz D, Weniger T, Urban B, Krüger F, Chen YG, 2013. Optically stimulated luminescence dating of young fluvial deposits of the Middle Elbe River Flood Plains using different age models. *Geochronometria* 41(1): 36–56, DOI 10.2478/s13386-013-0140-7.

Lanos P and Philippe A, 2015. Hierarchical Bayesian modeling for combining Dates in archaeological context. *Journal de la société française de statistique* 158(2): 72–88.

Li B, 2007. A note on estimating the error when subtracting background counts from weak OSL signals. *Ancient TL* 25(1): 9–14.

Li B, Jacobs Z, Roberts RG, Galbraith R and Peng, J, 2017. Variability in quartz OSL signals caused by measurement uncertainties: Problems and solutions. *Quaternary Geochronology* 41: 11–25, DOI 10.1016/j.quageo.2017.05.006.

Mercier N, Kreutzer S, Christophe C, Guérin G, Guibert P, Lahaye C, Lanos P, Philippe P and Tribolo C, 2016. Bayesian statistics in luminescence dating: The "baSAR"-model and its implementation in the R package "Luminescence". *Ancient TL* 34(2): 14–21.

Millard AR, 2004. *Taking Bayes beyond radiocarbon: Bayesian approaches to some other chronometric methods*. Lecture Notes in Statistics, Springer, London. DOI 10.1007/978-1-4471-0231-1_11.

Murray AS and Wintle AG, 2000. Luminescence dating of quartz using an improved single-aliquot regenerative-dose protocol. *Radiation Measurements* 32(1): 57–73, DOI 10.1016/S1350-4487(99)00253-X.

Olley JM, Roberts RG, Yoshida H and Bowler JM, 2006. Single-grain optical dating of grave-infill associated with human burials at Lake Mungo, Australia. *Quaternary Science Reviews* 25(19–20): 2469–2474, DOI 10.1016/j.quascirev.2005.07.022.

Peng J and Dong ZB, 2014. A simple Bayesian method for assessing the standard error of equivalent dose estimates. *Ancient TL* 32(2): 17–23.

Peng J, Dong ZB and Han FQ, 2016a. Application of slice sampling method for optimizing OSL age models used for equivalent dose determination. *Progress in Geography* 35(1): 78–88. (in Chinese)

Peng J, Dong ZB and Han FQ, 2016b. Optically stimulated luminescence dating of sandy deposits from Gulang county at the southern margin of the Tengger Desert, China. *Journal of Arid Land* 8(1): 1–12, DOI 10.1007/s40333-015-0137-6.

Peng J, Dong ZB, Han FQ, Long H and Liu XJ, 2013. R package numOSL: numeric routines for optically stimulated luminescence dating. *Ancient TL* 31(2): 41–48.

Peng J and Li B, 2017. Single-aliquot regenerative-dose (SAR) and standardised growth curve (SGC) equivalent dose determination in a batch model using the R package "numOSL". *Ancient TL* 35(2): 32–53.

Philippe A, Guérin G and Kreutzer S, 2019. BayLum - An R package for Bayesian analysis of OSL ages: An introduction. *Quaternary Geochronology* 49: 16–24, DOI 10.1016/j.quageo.2018.05.009.

Ramsey CB, 1995. Radiocarbon calibration and analysis of stratigraphy: the OxCal program. *Radiocarbon* 37(2): 425–430.

Ramsey C B, 2008. Deposition models for chronological records. *Quaternary Science Reviews* 27(1–2): 42–60, DOI 10.1016/j.quascirev.2007.01.019.

Rhodes EJ, Ramsey CB, Outram Z, Batt C, Willis L, Dockrill S and Bond J, 2003. Bayesian methods applied to the interpretation of multiple OSL dates: high precision sediment ages from Old Scatness Broch excavations, Shetland Isles. *Quaternary Science Reviews* 22(10–13): 1231–1244, DOI 10.1016/S0277-3791(03)00046-5.

Roberts RG, Galbraith RF, Yoshida H, Laslett GM and Olley JM, 2000. Distinguishing dose populations in sediment mixtures: a test of single-grain optical dating procedures using mixtures of laboratory-dosed quartz. *Radiation Measurements* 32(5–6): 459–465, DOI 10.1016/S1350-4487(00)00104-9.

Schmidt S, Tsukamoto S, Salomon E, Frechen M and Hetzel R, 2012. Optical dating of alluvial deposits at the orogenic front of the Andean Precordillera (Mendoza, Argentina). *Geochronometria* 39(1): 62–75, DOI 10.2478/s13386-011-0050-5.

Sivia DS, Burbidge C, Roberts RG and Bailey RM, 2004. A Bayesian approach to the evaluation of equivalent doses in sediment mixtures for luminescence dating. *AIP Conference Proceedings* 735(1): 305–311, DOI 10.1063/1.1835227.

Stan Development Team, 2018. *RStan: the R interface to Stan*. R package version 2.17.3.

Tamura T, Cunningham AC and Oliver TSN, 2019. Two-dimensional chronostratigraphic modelling of OSL ages from recent beach-ridge deposits, SE Australia. *Quaternary Geochronology* 49: 39–44, DOI 10.1016/j.quageo.2018.03.003.

Thomsen, K.J., Murray, A.S., Bøtter-Jensen, L., 2005. Sources of variability in OSL dose measurements using single grains of quartz. *Radiation Measurements* 39(1): 47–61, DOI 10.1016/j.radmeas.2004.01.039.

Zeeden C, Dietze M and Kreutzer S, 2018. Discriminating luminescence age uncertainty composition for a robust Bayesian modelling. *Quaternary Geochronology* 43: 30–39, DOI 10.1016/j.quageo.2017.10.001.

Zink A, 2013. A coarse Bayesian approach to evaluate luminescence ages. *Geochronometria* 40(2): 90–100, DOI 10.2478/s13386-013-0105-x.

Zink AJC, 2015. Bayesian analysis of luminescence measurements. *Radiation Measurements* 81: 71–77, DOI 10.1016/j.radmeas.2015.04.009.