

切片采样算法在光释光年龄模型参数估计中的应用

彭俊¹, 董治宝¹, 韩凤清²

(1. 中国科学院寒区旱区环境与工程研究所, 兰州 730000; 2. 中国科学院青海盐湖研究所, 西宁 810008)

摘要:光释光(OSL)年代学模型是基于数理统计学的一类概率密度模型,它根据特定的假设条件对样品等效剂量(De)分布进行数学解释,由此估计具有不同沉积历史或者能够代表样品实际埋藏年龄的 De 组分。年龄模型参数估计常通过极大似然估计(MLE)算法实现,本文尝试了切片采样算法在年龄模型参数优化中的应用。切片采样属于一种马尔科夫链蒙特卡罗采样(MCMC)算法,能根据测量数据与模型的联合似然函数进行随机采样,由此获得参数的采样分布。本文编写了实现年龄模型切片采样算法的应用程序,并使用模拟及实测 De 数据验证了该算法估计的可靠性。相对于MLE算法,MCMC算法具有对参数初值依赖性低、误差估计更准确的特点,切片采样算法提供了实现释光年龄模型参数估计的一种新方法。

关键词:OSL测年;等效剂量;年龄模型;参数优化;随机采样

1 引言

单片再生剂量法(SAR)(Murray et al, 2000)对样品不同单片的测量可获得一系列 De 值及其误差估计。自然样品的 De 分布常表现出一定的分散性,这与样品颗粒的侵蚀、搬运、沉积历史与埋藏环境有关。导致样品 De 分散的不确定性因素包括内部和外部误差(Duller, 2008),前者主要来源于实验操作、数据分析处理及仪器精度,包括光子计数统计不确定性、曲线拟合误差、仪器可重复率、人工贝塔源辐射剂量率的均一程度、实验室与自然条件下加热与辐射过程的差异、样品不同颗粒释光特性的差异等;后者指样品在搬运与埋藏过程中不同颗粒吸收辐射剂量的变化与不确定性,如埋藏前晒退程度不同造成不同颗粒残余剂量的差异、埋藏环境中不同颗粒贝塔剂量率的空间差异、不同埋藏历史与年龄的颗粒发生混合等(Thomsen et al, 2007)。这些都是可能造成自然沉积样品(尤其是水成沉积物) De 值异常分散的原因,即使测量采用的不是单颗粒

技术而是包含几十到上百粒样品的多颗粒单片(Rodnight et al, 2006; Schmidt et al, 2012; Keen-Zeibert et al, 2013)。对于 De 分布分散的样品,一般不能通过均值计算其埋藏剂量(Galbraith et al, 2012)。

有限组分混合年龄模型(FMM)(Galbraith et al, 1990; Roberts et al, 2000)和最小年龄模型(MAM)(Van der Touw et al, 1997; Galbraith et al, 1999)是两种应用广泛的统计学年齡模型,其模型参数常使用MLE算法来优化。但MLE算法的缺点在于优化结果对参数的初值敏感,模型容易陷入局部最优解。对于多参数模型或高噪声观测数据,MLE算法需比较以不同初值初始化的优化结果才能确定模型的全局最优解。此外,MLE算法不能直接估计参数的标准误差,参数误差常通过费歇尔信息矩阵来估计。对于样本量无限大的样本数据,假定模型的对数似然函数为二次型形式,则模型参数的方差—协方差矩阵等价于费歇尔信息矩阵的逆矩阵(Neale et al, 1997),对该逆矩阵主对角元素取平方根即可得到参数标准误差的近似估计。然而一般统计模

收稿日期:2015-04;修订日期:2015-08。

基金项目:国家重点基础研究发展计划(973计划)项目(2013CB956000,2012CB426501) [Foundation: National Basic Research Program of China (973 Program), No.2013CB956000, No.2012CB426501]。

作者简介:彭俊(1987-),男,湖北红安人,博士生,主要从事释光年代学与气候变化研究,E-mail: pengjun10@mails.ucas.ac.cn。

引用格式:彭俊,董治宝,韩凤清. 2016. 切片采样算法在光释光年龄模型参数估计中的应用[J]. 地理科学进展, 35(1): 78-88. [Peng J, Dong Z B, Han F Q. 2016. Application of slice sampling method for optimizing OSL age models used for equivalent dose determination[J]. Progress in Geography, 35(1): 78-88.]. DOI: 10.18306/dlkxjz.2016.01.009

型的对数似然函数(包括本文分析的年龄模型)很难满足二次型形式,造成基于信息矩阵的参数误差估计不准确。近年来,随机采样算法在释光年代学数据分析中得到广泛应用,Rhodes等(2003)使用贝叶斯方法将样品地层序列与释光年代数据并入沉积模型,并使用MCMC算法对年代序列进行了校正;Sivia等(2004)应用MCMC算法实现了基于下限 De 误差的混合分布模型参数估计;Duller(2007)使用蒙特卡罗算法模拟了 De 计算的误差传递过程;Hunt-riss(2008)使用MCMC采样技术分析了联合线性剂量模型 De 分布与多温度预热评估模型初始温度分布;Zink(2013)使用MCMC技术分析了样品沉积过程中多种因素对埋藏剂量分布的影响;Peng等(2014)使用MCMC算法改进了SAR法 De 误差估计方法,提高了 De 估计的精度;彭俊等(2015)使用蒙特卡罗算法模拟了剂量率计算过程中多变量耦合误差的传递过程。本文简要介绍了一种MCMC算法(切片采样)的随机采样原理,尝试应用该算法估计释光年龄模型的参数及误差,旨在克服MLE算法对模型参数初值敏感及误差估计不准确的缺点,编写了执行MCMC随机采样算法的程序软件,并根据模拟及实测 De 数据验证了该算法参数估计的可靠性。

2 模型简介

FMM模型常用来估计呈多峰分布的 De 数据不同组分的特征 De 值,微尺度上辐射剂量率的不均一是造成 De 分布呈多峰现象的原因之一,临近放射性活动强的矿物(如钾长石和锆石)的颗粒剂量率高于平均水平,而接近低放射性活动物质(如碳酸钙结核)的颗粒剂量率低于平均水平,由此造成同一沉积层位的样品不同颗粒吸收的累积辐射剂量的空间差异。此外,生物沿垂直地层上下移动造成具有不同年龄的颗粒发生混合(Bateman et al, 2007),也可能造成 De 数据呈多峰分布。假定单组分 De 值服从正态分布(Galbraith, 2003),则 k 组分FMM模型的对数联合似然函数可表示为:

$$\begin{aligned} x_i &= \frac{s(z_i)}{z_i}; y_i = \ln(z_i); \\ \sum_{j=1}^k p_j &= 1; f_{ij} = \frac{1}{\sqrt{2\pi(x_i^2 + \sigma^2)}} e^{-\frac{(y_i - u_j)^2}{2(x_i^2 + \sigma^2)}}; \\ L_{FMM} &= \sum_{i=1}^n \ln(p_1 f_{i1} + p_2 f_{i2} + \cdots + p_k f_{ik}) \end{aligned} \quad (1)$$

式(1)中: z_i 与 $s(z_i)$ 分别表示第 i 个 De 值及其测量误差; y_i 与 x_i 分别表示第 i 个自然对数 De 值及其相对误差; p_j 和 μ_j 分别表示第 j 个 De 组分的比重及特征 De 值(自然对数尺度); f_{ij} 表示第 i 个 De 值对第 j 个 De 组分似然函数的贡献; σ 为 De 分布的已知附加误差(相对误差尺度),表示SAR法 De 误差估计中被忽略的误差来源,包括仪器可重复性误差、非泊松分布的光子计数统计误差、热转移效应误差以及人工贝塔源辐射剂量率误差等; L_{FMM} 表示FMM模型的对数似然值; n 表示 De 值的数目。模型组分数 k 可通过贝叶斯信息准则(Schwarz, 1978)估计,关于贝叶斯信息准则在模型选择中的应用可参考Peng等(2014)。

在此有必要解释一下模型参数估计之前对观测数据(De 及其误差)作自然对数转换的必要性。因为通常SAR法估计的 De 绝对误差随 De 增加而增大,造成实测 De 数据服从概率密度曲线呈正偏态的对数正态分布,对观测数据求自然对数(出现负 De 值的年轻样品除外)可将其转换为正态 De 分布(对数正态分布变量的对数值服从正态分布),这有利于消除正偏态数据结构对年龄模型优化结果可靠性的影响(因为年龄模型要求单组份 De 服从正态分布)。而将 De 的绝对标准误差转化为相对标准误差的原因主要有两点:(1)相对标准误差受 De 值变化的影响较小(Arnold et al, 2009),为比较同一样品不同测片 De 误差及不同样品不同测片 De 误差提供了更合理的标准;(2)对于同一变量,其相对误差近似等于其自然对数值的绝对误差(Galbraith et al, 2012),因此在对 De 值取自然对数后,只有相应地将 De 值的绝对误差变为相对误差才能维持原始数据中 De 值与误差值之间的幅度关系。

样品颗粒埋藏前晒退不完全也是造成 De 分布分散的原因之一,用于晒退不完全沉积样品 De 估计的统计方法有多种,Olley等(1998)使用 De 分布最低5%分位数内的单片 De 均值来估计样品 De ;Lepper等(2002)使用正态分布函数拟合 De 分布直方图的前缘区域,将拟合获得的正态分布函数中二阶导数为零的点确定为样品 De ;Zhang等(2003)通过比较经感量校正的再生OSL信号与自然OSL信号的相对标准偏差来选择晒退完全的单片 De ,并将其均值作为样品 De ;Thomsen等(2007)将呈升序排列的 De 数据“内部”与“外部”测量误差比值为整数1的单片 De 均值作为样品 De ;Pietsch(2009)使用正态分布函数拟合 De 分布多峰概率密度曲线的最前

峰,将拟合得到的正态分布函数均值作为样品 De 。但这些选取晒退完全单片 De 组分的方法都属于经验性方法,缺乏严格的数理统计学理论作为依据,估计的样品 De 可靠性在很大程度上依赖于分析所用的样品数据。相比之下, MAM 模型是应用最广泛的分析晒退不完全 De 分布的统计学模型,它的建立以严格数理统计学准则为基础。该模型分3参数(MAM_3)与4参数(MAM_4)两种类型, MAM_4 模型对 De 数据内在结构(如 De 分布的离散度、 De 值的数目、 De 数据的相对误差等)的要求较高,数值稳定性差,因而应用较少。相比之下, MAM_3 普适性强,数值稳定性高,被广泛用于不同类型沉积样品的 De 估计(Schmidt et al, 2012; Kunz et al, 2013; Ou et al, 2015)。假设样品颗粒因埋藏前被晒退到不同程度而造成晒退完全与不完全的颗粒共存,由此样品 De 分布可看作由完全晒退组分和未完全晒退组分按一定比例形成的混合概率密度分布函数,则 MAM_3 模型的对数联合似然函数可表示为:

$$\gamma_0 = \frac{\gamma}{\sigma^2} + \frac{y_i}{x_i^2}; \sigma_0 = \frac{1}{\sqrt{\frac{1}{\sigma^2} + \frac{1}{x_i^2}}};$$

$$f_{i1} = \frac{1}{\sqrt{2\pi x_i^2}} e^{-\frac{(y_i - \gamma)^2}{2x_i^2}}; f_{i2} = \frac{2[1 - \Phi(\frac{\gamma - \gamma_0}{\sigma_0})]}{\sqrt{2\pi(x_i^2 + \sigma^2)}} e^{-\frac{(y_i - \gamma)^2}{2(x_i^2 + \sigma^2)}}; \quad (2)$$

$$L_{MAM3} = \sum_{i=1}^n \ln[p f_{i1} + (1-p) f_{i2}]$$

式(2)中: γ_0 与 σ_0 为模型优化过程中环节的代用参数,并无实际物理意义; y_i 与 x_i 分别表示第 i 个对数 De 值及其相对误差(与式(1)相同); f_{i1} 与 f_{i2} 分别第 i 个 De 值对晒退完全组分与非完全组分似然函数的贡献;模型的3个待估参数为 p 、 γ 、 σ , 其中 p 表示晒退完全 De 组分的比重, γ 代表晒退完全 De 组的特征 De 值(自然对数尺度), σ 为表征未完全晒退组分 De 离散程度的附加误差(相对误差尺度); $\Phi(x)$ 表示标准正态分布累积函数, L_{MAM3} 表示 MAM_3 模型的对数似然值。

3 切片采样算法原理及程序实现

Metropolis-Hasting 和 Gibbs 是常见的两种随机采样算法,这两种算法对概率密度模型的适应性因具体模型而异。Metropolis-Hasting 算法需先指定一个便于随机数生成的“建议”分布函数,然后以

“接受—拒绝”方式决定是否接受新生成的随机数。“建议”分布与目标分布函数越相似,采样效果越好。但对于复杂分布函数,如何选择“建议”分布函数并非易事,它直接影响到目标分布函数样本的质量。在 Gibbs 采样中,模型不同参数通过条件概率密度分布发生联系,随机变量出现(生成)的概率由其满条件分布和其余变量的当前值共同决定,通过依次轮流的方式实现多参数样本更新(Gelman et al, 2013; Lunn et al, 2013)。Gibbs 算法的缺点在于需先确定参数的满条件分布,满条件分布为标准分布函数(如指数、正态分布等)的参数可直接采样,但满条件分布复杂的变量需使用其他算法(如 Metropolis-Hasting、切片采样算法等)获得其采样分布。

切片采样(Neal, 2003)为一种广义的 Gibbs 采样算法,属于附加变量采样技术。Neal(2003)指出,单变量概率函数的采样分布可通过对其概率密度曲线下区域对应变量值的随机均匀采样获得。切片采样算法在生成密度函数 $f(x)$ 的随机变量 x_1 之前,需引入附加变量 y , 其生成随机变量的方法通过如下迭代步骤实现:①给定初值 x_0 , 然后从区间 $(0, f(x_0))$ 中随机均匀地生成附加变量 y , 切片 S 对应的横轴 x 值集合为 $\{x; f(x) > y\}$; ②确定以 x_0 为扩张基点的采样区间 $I = (L, R)$, 该区间应尽可能的覆盖水平切片 S 的绝大部分; ③从采样区间 I 中随机均匀地生成新变量 x_1 。图1为单变量切片采样算法示意图,图中由均匀分布随机变量 y 确定的水平切片 S 被概率密度函数 $f(x)$ 分割为小切片 S_1 、 S_2 、 S_3 , 与切片 S 对应的横轴所有 x 值满足 $f(x) > y$ 。以 x_0 为基点的样本区间以 w 为步长持续增加宽度,直到该样本区间的两端扩张到切片 S 以外的区域,然后从扩张后的样本区间 $I = (L, R)$ 中随机均匀地生成新变量 x_1 。

切片采样算法最棘手环节在于如何确定采样区间 I 的边界使其尽可能的覆盖切片 S 。若能确定

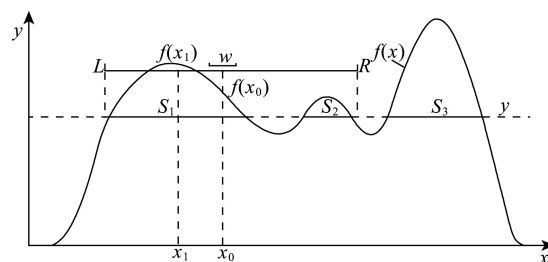


图1 单变量概率密度函数的切片采样算法实现示意图
Fig.1 Single-variable Slice sampling by simulating a sample from the region under its density function $f(x)$

概率密度函数 $f(x)$ 反函数 $f^{-1}(x)$ 的解析式,便能精确界定采样区间 I 的边界。但很多复杂概率密度模型不能被反解,且函数 $f(x)$ 的起伏变化容易对水平切片 S 形成切割作用,使得区间 I 呈现多重边界因而很难逐一界定。为此,Neal(2003)提出了名为“Stepping out”和“Shrinkage”的方法来确定区间 I 边界。首先以 x_0 为基点定义宽度为 w 的微型水平区间 I ,然后以 w 为步长重复扩张该区间,直到区间 I 的两端超出水平切片 S 的范围,即 $f(x) \leq y$ 。由于采样区间 $I=(L,R)$ 对应的横轴 x 值不一定全部满足 $f(x) > y$,所以若新生成的随机数 x_1 位于切片外部,则需要收缩区间 I 两端,然后再次生成 x_1 ,重复此过程直到生成的 x_1 满足 $f(x_1) > y$ 为止,随着不满足条件的随机变量不断被拒绝及区间 I 的逐步收缩,生成符合条件的随机数 x_1 的概率逐渐增加。

通过类似于Gibbs算法对不同变量轮流依次采样的方式(表1),将每个参数的分布函数当作条件性依赖于其他参数当前值的单变量概率模型,则上述单变量切片采样算法可直接推广应用于多参数概率密度模型采样。本文编写了进行MAM与FMM模型切片采样的开源R软件程序(Peng et al, 2013; 彭俊等, 2016) (<http://CRAN.R-project.org/package=numOSL>)。由于MCMC采样过程消耗内存且费时,为提高运算效率,本文核心程序使用Fortran语言编写,通过接口方式在R软件环境下调用。获得模型参数分布之后,便可根据中心极限定理来估计参数的期望值,并计算标准偏差、众数、百分位数等统计量。需要注意的是,由于模型分析时观测数据(De)被转换到自然对数尺度,因此算法运行结束后

需将 De 的采样分布转化到非自然对数尺度才能进行样本统计分析。此外,考虑到交换FMM模型不同组分参数位置不影响模型的似然函数值,例如,将FMM第1组分参数(p_1, μ_1)与第2组分参数(p_2, μ_2)交换位置并不影响对数似然函数 L_{FMM} 的值,混合模型组分参数的这种特性称为“标签置换”,该效应可能会引起随机参数在样本链中的位置发生置换变动。本程序通过对FMM模型生成的随机样本作 $\mu_1 < \mu_2 < \dots < \mu_k$ 的定向排序来减小“标签置换”效应对参数估计结果的影响。

4 基于模拟及实测 De 数据的算法验证

SAR测量的 De 值的绝对误差与其大小有关,较大 De 值的误差也较大(Galbraith et al, 2012)。然而, De 值的相对误差可认为独立于 De 值变化(Arnold et al, 2009)。本文以一组实测 De 数据的误差结构为参照,从参数值已知的年龄模型中生成随机 De 分布,旨在使模拟生成的随机 De 数据的误差结构更接近实测SAR De 的误差结构。所用的实测 De 数据的相对误差近似服从均值为7.5%,标准偏差为5%的对数正态分布(Arnold et al, 2009)。随机 De 数据按如下方式生成:①确定模型参数及随机 De 值的初始状态 y_0 ;②生成服从上述对数正态分布的 De 相对误差 x ;③使用单变量切片采样算法从参数已知的年龄模型中生成随机 De 值 y (对数尺度), y 条件地依赖于 y_0 与 x ;④将对数尺度 De 值 y 及其相对误差 x 转换为自然尺度 De 值 z 及绝对误差 $s(z)$,即, $z = \exp(y)$, $s(z) = \exp(y) \cdot x$ 。本文根据上述方法使用列于表2-3中的已知参数从不同年龄模型(MAM₃和FMM₃)中模拟生成随机 De 数据,对每个模型生成两组离散程度不同的 De 分布,分别命名为 S_1 -MAM₃、 S_2 -MAM₃、 S_1 -FMM₃、 S_2 -FMM₃。对于每组已知参数,先模拟生成10000个随机 De 值(含 De 误差),然后舍弃前7500个数据以降低 De 初始状态 y_0 对随机 De 分布的影响(“Burn-in”程序),最后从剩余的2500个数据中每10个取1个以降低随机 De 值间的自相关程度(“Thinning”程序),每组得到250个 De 数据,由此获得的4组模拟样品 De 分布见图2。

本文在应用切片采样算法估计模拟与实测样品 De 数据的模型参数时,MCMC算法采样总数均为50000次,舍弃前10000个样本值(“Burn-in”数为

表1 多变量切片采样算法流程

Tab.1 A general procedure for applying the Slice sampling method to sampling from a multivariable model

输入: 概率模型, 迭代总次数 $nsim$, 初始参数状态 X_0 , 模型参数总数 N ;
输出: 随机采样样本及参数估计结果
1. 分配用于随机样本储存的数据矩阵(chain): $chain = \text{matrix}(\text{nrow} = nsim, \text{ncol} = N)$;
2. 通过如下迭代方式对参数轮流更新:
do $i=1, nsim$
do $j=1, N$
根据切片采样算法以条件概率 $P(X_j/X_0[-j])$ 生成随机变量 X_j
$X_0[j] = X_j$
$chain[i,j] = X_j$
end do
end do
3. 根据随机样本数据chain进行参数估计

10000),然后从剩余样本中每5次迭代取一个样本(“Thinning”数为5),剩余8000个样本用于模型参数估计。模拟样品 De 数据的 MAM_3 与 FMM_3 模型参数估计结果分别见表2-3,结果表明根据MCMC与MLE算法估计的模型参数值均与实际值是一致的,说明基于切片采样算法的年龄模型参数估计结果是可靠的。图3为使用参数 $p=0.1$ 、 $\gamma=10$ 、 $\sigma=0.25$ 生成的随机 De 分布(模拟样品 S_1 - MAM_3)的 MAM_3 模型MCMC算法采样结果,图中第1列为参数的密度分布图,第2列为随机样本随迭代次数变化图,第3列为参数自相关系数变化图。从图中可以看出,随机样本具有良好的混合特性,参数的自相关程度都很低,表明MCMC采样算法收敛到了模型参数的

期望分布。此外,图3中 MAM_3 模型参数 p 的采样分布呈正偏态,MLE算法估计的 p 值为 0.08 ± 0.09 ,

表2 根据已知模型参数生成的 De 数据的 MAM_3 参数估计
Tab.2 Estimates of MAM_3 obtained from De distributions simulated using known parameters

样品名称	实际参数	MCMC算法参数估计	MLE算法参数估计
S_1 - MAM_3	$p=0.1$	$p=0.09\pm0.06$ (0.005, 0.22)	$p=0.08\pm0.09$ (-0.10, 0.25)
	$\gamma=10$	$\gamma=9.92\pm0.13$ (9.64, 10.17)	$\gamma=9.92\pm0.19$ (9.54, 10.29)
	$\sigma=0.25$	$\sigma=0.25\pm0.01$ (0.23, 0.28)	$\sigma=0.25\pm0.01$ (0.22, 0.28)
S_2 - MAM_3	$p=0.3$	$p=0.20\pm0.04$ (0.12, 0.28)	$p=0.20\pm0.04$ (0.12, 0.28)
	$\gamma=10$	$\gamma=9.82\pm0.08$ (9.66, 9.98)	$\gamma=9.83\pm0.08$ (9.68, 9.98)
	$\sigma=0.5$	$\sigma=0.54\pm0.03$ (0.48, 0.60)	$\sigma=0.54\pm0.03$ (0.48, 0.59)

注:括号内数值表示参数的95%置信区间,下表同。

表3 根据已知模型参数生成的 De 数据的 FMM_3 参数估计

Tab.3 Estimates of FMM_3 obtained from De distributions simulated using known parameters

样品名称	实际参数	MCMC算法参数估计	MLE算法参数估计
S_1 - FMM_3	$p_1=0.1; \mu_1=10$	$p_1=0.12\pm0.02$ (0.08, 0.17); $\mu_1=10.06\pm0.11$ (9.85, 10.28)	$p_1=0.12\pm0.02$ (0.08, 0.16); $\mu_1=10.04\pm0.11$ (9.83, 10.26)
	$p_2=0.5; \mu_2=15$	$p_2=0.47\pm0.03$ (0.40, 0.53); $\mu_2=15.08\pm0.07$ (14.94, 15.23)	$p_2=0.47\pm0.03$ (0.41, 0.54); $\mu_2=15.08\pm0.07$ (14.94, 15.22)
	$p_3=0.4; \mu_3=20$	$p_3=0.41\pm0.03$ (0.35, 0.47); $\mu_3=20.10\pm0.11$ (19.89, 20.32)	$p_3=0.41\pm0.03$ (0.34, 0.47); $\mu_3=20.10\pm0.11$ (19.89, 20.32)
S_2 - FMM_3	$p_1=0.3; \mu_1=10$	$p_1=0.27\pm0.03$ (0.22, 0.33); $\mu_1=10.07\pm0.06$ (9.97, 10.18)	$p_1=0.27\pm0.03$ (0.21, 0.33); $\mu_1=10.07\pm0.06$ (9.96, 10.18)
	$p_2=0.35; \mu_2=20$	$p_2=0.41\pm0.03$ (0.35, 0.47); $\mu_2=20.12\pm0.09$ (19.94, 20.30)	$p_2=0.41\pm0.03$ (0.35, 0.48); $\mu_2=20.12\pm0.09$ (19.94, 20.30)
	$p_3=0.35; \mu_3=30$	$p_3=0.32\pm0.03$ (0.26, 0.38); $\mu_3=29.95\pm0.16$ (29.64, 30.26)	$p_3=0.32\pm0.03$ (0.26, 0.38); $\mu_3=29.95\pm0.16$ (29.64, 30.26)

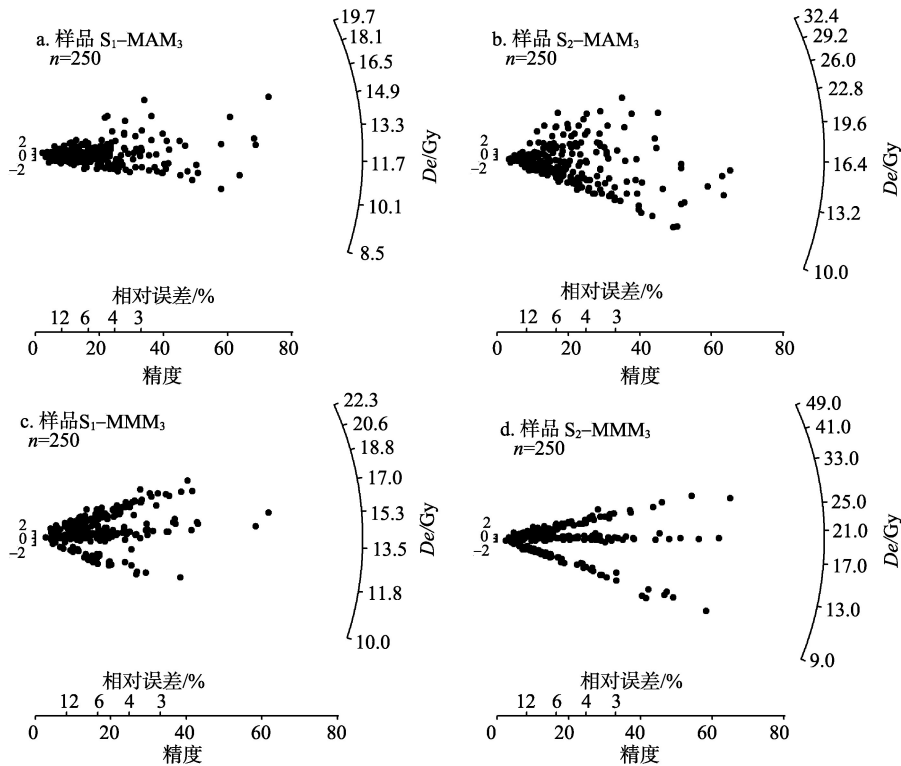
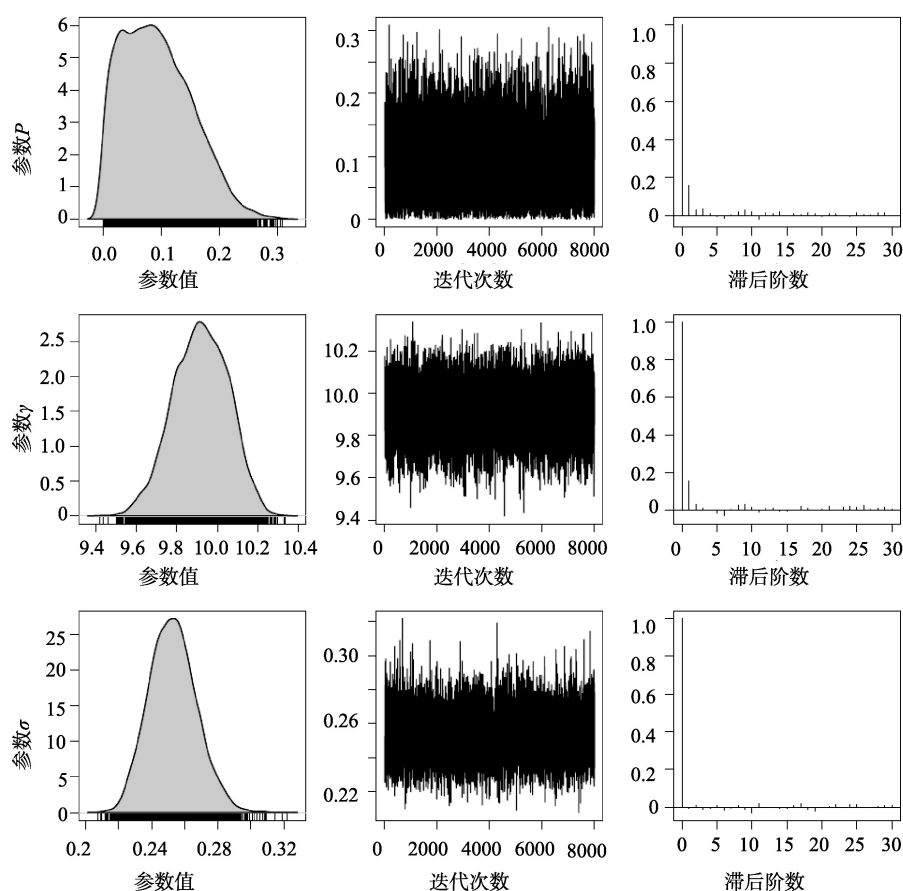


图2 根据已知模型参数模拟生成样品的 De 分布

Fig.2 De distributions for simulated samples using age models with known parameters

图3 模拟样品S₁-MAM₃ De数据的MAM₃模型MCMC算法采样结果Fig.3 MCMC output of MAM₃ using simulated *De* values from sample S₁-MAM₃

由此根据正态近似法(即 $value \pm 1.96\sigma$, $value$ 表示参数点估计, σ 为参数标准偏差)计算的参数 p 的95%置信区间为 $(-0.10, 0.25)$, 该区间下限小于零而不符合实际, 而根据MCMC算法获得的随机样本分布计算的 p 值95%置信区间为 $(0.005, 0.22)$, 可见当参数呈偏态分布时, 由正态近似法获得的对称性置信区间不可靠(Galbraith et al, 2012)。

需要注意的是, 验证切片采样算法的实用性时, 需考虑到模拟数据与实测数据之间的差异。因为尽管以上模拟样品 *De* 数据的随机生成过程考虑了实测样品 *De* 分布的误差结构, 但仅能代表理想状态下的 *De* 分布。自然沉积环境中样品不同颗粒辐射剂量率的空间差异、晒退程度差异、不同沉积层位颗粒混合等增加了实测 *De* 数据的噪声水平, 造成实测 *De* 分布复杂而分散。此外, 上述每组模拟数据多达250个 *De* 值因而包含了丰富的信息量, 这有利于提高模型采样过程的稳定性, 但实验室实测样品单片 *De* 数通常在50以内(一般不超过

100)。为此, 本文使用两个实测水成沉积样品(142/SC08-03 和 AL3)的 *De* 数据对MCMC算法可靠性作进一步验证(图4)。样品142/SC08-03为来自南非东部内陆的河流沉积物, 共测量65个 *De* 值(Keen-Zebert et al, 2013), 样品AL3为来自南美安第斯山脉的冲积扇沉积物, 共测量84个 *De* 值(Schmidt et al, 2012)。本文在实测数据相对误差的基础上并入10%的附加误差来表示 *De* 误差估计中被忽略的误差来源, 以免 *De* 误差被低估(Galbraith et al, 2005)。由于年龄模型对含异常 *De* 值的观测数据敏感, 本文在数据分析之前移除了实测 *De* 数据AL3中的一个异常值(如图4b虚线方框所示), 然后使用MCMC和MLE算法分析了两组数据并对结果进行了比较(表4-5)。对比结果表明, 两种算法的参数估计结果很一致。图5为样品142/SC08-03的MAM₃模型MCMC采样结果, 参数 p 同样呈明显正偏态分布, MLE算法估计的 p 值为 0.14 ± 0.08 , 由此根据正态近似法估计的95%置信区间为 $(-0.02,$

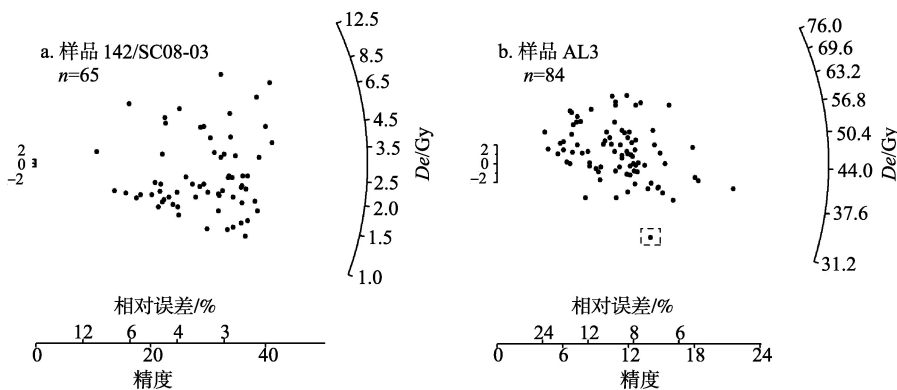


图4 实测样品 142/SC08-03 与 AL3 的 De 分布
Fig.4 De distributions for measured samples 142/SC08-03 and AL3

0.30), 而根据 MCMC 算法采样分布计算的 p 值 95% 置信区间为 (0.01, 0.29), 后者更合理。样品 AL3 的 MAM₃ 模型 MCMC 分析结果中参数 p 结果类似, MLE 算法结合正态近似法估计的 95% 置信区间为

(-0.21, 0.60), MCMC 算法估计的 p 值 95% 置信区间为 (0.01, 0.51)。

5 MCMC 算法的收敛诊断

通过给定参数的初始状态(类似于优化问题中的参数初值), MCMC 算法以随机游走的方式从概率密度模型采样, 采样器能否经过多次迭代后“忘记”初始状态并收敛到模型参数的平稳分布是进行参数估计的前提。用于诊断马氏链是否收敛的方法有多种(Plummer et al, 2006; Lunn et al, 2013), 最简单的方法是观察随机样本值随迭代次数的变化, 均匀混合的随机样本可在参数状态空间上下自由移动(如图 3 与图 5 中第 2 列)。若随机样本随迭代次数呈阶梯步状式的大幅震荡变化, 则说明样本混合不均匀, 算法可能未收敛。通过随机样本的自相关系数也可以检验 MCMC 算法的收敛性: 样本的

表 4 实测 De 数据的 MAM₃ 模型参数估计
Tab.4 Estimates of MAM₃ obtained using measured De distributions

样品名称	MCMC 算法参数估计	MLE 算法参数估计
142/SC08-03	$p=0.14\pm0.07$ (0.01, 0.29)	$p=0.14\pm0.08$ (-0.02, 0.30)
	$\gamma=1.06\pm0.05$ (0.95, 1.15)	$\gamma=1.07\pm0.05$ (0.97, 1.18)
	$\sigma=1.18\pm0.12$ (0.97, 1.44)	$\sigma=1.15\pm0.11$ (0.92, 1.37)
	$L=-59.34$	$L=-59.26$
AL3	$p=0.22\pm0.14$ (0.01, 0.51)	$p=0.20\pm0.21$ (-0.21, 0.60)
	$\gamma=40.56\pm1.90$ (36.78, 44.10)	$\gamma=40.49\pm2.58$ (35.44, 45.54)
	$\sigma=0.41\pm0.05$ (0.32, 0.52)	$\sigma=0.39\pm0.05$ (0.30, 0.48)
	$L=-9.53$	$L=-9.45$

注: L 表示算法计算的模型对数极大似然值, 下表同。

表 5 实测 De 数据的 FMM 模型参数估计
Tab.5 Estimates of FMM obtained using measured De distributions

样品名称	MCMC 算法参数估计	MLE 算法参数估计
142/SC08-03	$p_1=0.30\pm0.05$ (0.21, 0.40); $\mu_1=1.15\pm0.03$ (1.09, 1.21)	$p_1=0.30\pm0.06$ (0.19, 0.42); $\mu_1=1.14\pm0.03$ (1.08, 1.20)
	$p_2=0.35\pm0.05$ (0.25, 0.45); $\mu_2=1.94\pm0.05$ (1.86, 2.03)	$p_2=0.37\pm0.06$ (0.25, 0.49); $\mu_2=1.94\pm0.04$ (1.85, 2.02)
	$p_3=0.22\pm0.05$ (0.13, 0.31); $\mu_3=4.22\pm0.14$ (3.95, 4.51)	$p_3=0.20\pm0.05$ (0.10, 0.30); $\mu_3=4.19\pm0.14$ (3.93, 4.46);
	$p_4=0.13\pm0.04$ (0.06, 0.23); $\mu_4=9.66\pm0.47$ (8.82, 10.65)	$p_4=0.12\pm0.04$ (0.04, 0.20); $\mu_4=9.54\pm0.42$ (8.72, 10.35)
	$L=-97.82$	$L=-97.62$
AL3	$p_1=0.39\pm0.11$ (0.18, 0.61); $\mu_1=41.31\pm1.73$ (37.69, 44.55)	$p_1=0.39\pm0.13$ (0.14, 0.65); $\mu_1=41.23\pm1.77$ (37.77, 44.69)
	$p_2=0.38\pm0.10$ (0.17, 0.58); $\mu_2=52.91\pm3.42$ (47.08, 60.79)	$p_2=0.40\pm0.12$ (0.15, 0.64); $\mu_2=53.01\pm3.08$ (46.97, 59.05)
	$p_3=0.23\pm0.06$ (0.12, 0.36); $\mu_3=79.01\pm4.41$ (70.86, 88.12)	$p_3=0.21\pm0.06$ (0.10, 0.32); $\mu_3=79.72\pm4.14$ (71.60, 87.83)
	$L=-6.83$	$L=-6.74$

注: 应用贝叶斯信息准则(Schwarz, 1978)估计的样品 142/SC08-03 与 AL3 的最佳 De 组分数分别为 4 和 3, 对应 FMM 模型分别为 FMM₄, FMM₃。

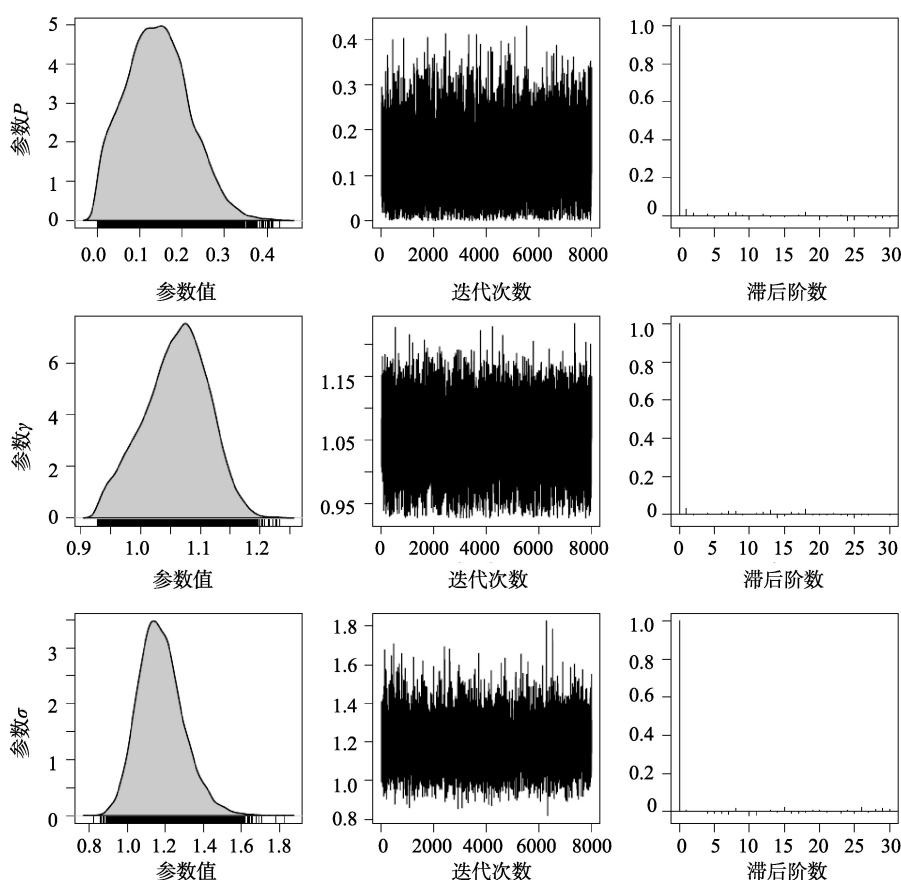


图5 实测样品 142/SC08-03 De 数据的 MAM_3 模型 MCMC 算法采样结果

Fig.5 MCMC output of MAM_3 using measured De values from sample 142/SC08-03

自相关程度越低(如图3和图5中第3列),说明随机样本混合特性越好,算法收敛越快。通常在使用随机样本进行统计推断之前,先要对样本数据进行“修剪”,这是因为随机样本的初始部分可能未收敛到参数的平稳分布而需将其删除(称为“Burn-in”程序),而为降低随机序列的自相关程度需从每数个样本中取1个(称为“Thinning”程序)。若“修剪”后随机样本的混合特性仍很差或自相关度仍很高,则说明算法未收敛。

Gelman-Rubin 收敛诊断法 (Gelman et al, 1992) 是应用最广泛的马氏链收敛诊断法,它对同一组观测数据使用不同初始状态随机采样由此获得多组平行马氏链,然后根据收缩因子(Shrink factor)来测量链内与链间方差的差异(类似于方差分析),若收缩因子小于等于1则说明算法收敛。本文设置5组不同初始状态,应用 MAM_3 模型分析了实测样品 AL3 的 De 数据,并使用 Gelman-Rubin 收敛诊断 R 软件程序包 coda(Plummer et al, 2006)分析了根据切

片采样法生成的5组平行马氏链,结果见图6。在最初的300次迭代中,各参数(p 、 γ 、 σ)的收缩因子都出现大幅度震荡,但均在1000次迭代内降到1以下,说明随机生成的马氏链收敛到了 MAM_3 模型的平稳分布。

6 讨论与结论

在本文所有 De 数据(包括模拟和实测样品)的 MCMC 算法分析中,没有出现必须调整参数初始状态才能顺利完成随机采样的情况,而是一直使用程序默认设置的初始参数。此外,应用差异很大的初始状态于同一组 De 数据获得的马氏链的参数估计结果完全一致,表明本文所采用的切片采样器对参数的初始状态不敏感,能够在有限次迭代内收敛到模型的平稳分布(全局最优解)。而 MLE 算法在不同初始化条件下的参数估计不尽相同, Rex Galbraith 提出使用观测数据范围内等距间隔的 De 分

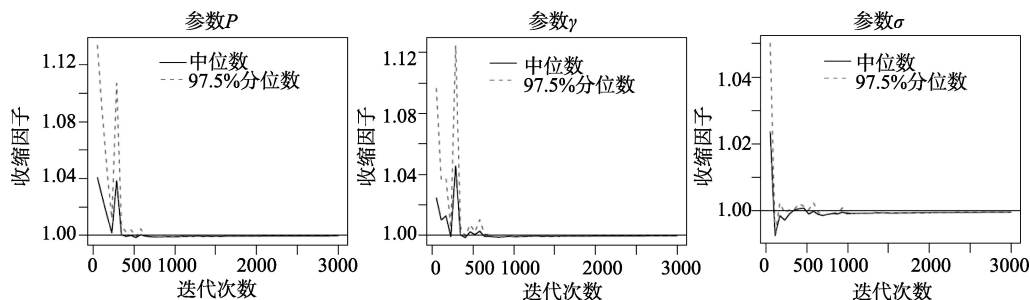


图6 应用MAM₃模型分析样品AL3 *De*数据的Gelman-Rubin收敛诊断结果

Fig.6 Gelman-Rubin diagnostic plots for MCMC output of MAM₃ using *De* values of sample AL3

位值作为FMM模型特征 De 初值,这种初始化方式在大多数情况下均有效,但对于高噪声 De 数据,其优化参数可能仅收敛到局部极值,因此应用MLE算法时有必要使用基于试错法的高强度、多初值反复优化的方式来确定模型的全局最优解。本文建议通过对同一组 De 数据进行不同算法结果的比较分析来确定合理的参数估计,以免造成对 De 分布的错误解释。

模拟样品 De 数据的分析结果表明切片采样算法的参数估计与模型实际参数是一致的,而两种不同算法对实测样品 De 数据的分析进一步验证了MCMC算法估计结果的可靠性。MLE算法常使用费歇尔信息矩阵近似代替方差—协方差矩阵来估计模型参数的误差,信息矩阵可通过对参数求偏导获得,复杂模型的信息矩阵常采用有限差分逼近来计算,逆信息矩阵主对角元素的平方根可近似作为参数标准误差。但基于信息矩阵的参数不确定性估计要求模型似然函数为二次型形式(Neale et al, 1997),而一般似然模型并不满足该要求。而在MCMC算法中,直接可靠的误差估计可通过计算采样分布的标准偏差实现(Galbraith et al, 2012)。此外,根据中心极限定理,当观测变量的采样分布趋于服从正态分布时,根据正态分布的对称性可使用正态近似法计算参数的置信区间,例如,假定参数 $value$ 的标准误差为 σ ,则对应的95%置信区间的近似估计为 $(value-1.96\sigma, value+1.96\sigma)$,但若参数的采样分布呈一定的偏度(如图3、图5中MAM₃模型的参数 p),则通过正态近似法计算的置信区间不可靠。而对于MCMC算法,参数的置信区间可直接根据采样分布计算,显著性水平 α 对应的 $100(1-\alpha)\%$ 置信区间由采样分布的 $(\alpha/2)\%$ 和 $(1-\alpha/2)\%$ 分位数决定,例如,显著性水平为0.05时参数 p 的95%置信区间介于 p 样本分布的2.5%与97.5%分位数之间。

必须指出,正如所有的MCMC算法一样,本文采用的切片采样算法也有自身缺点。模型参数的采样分布通常需要由数万次迭代计算(本文程序默认迭代次数为5万次)获得,随模型复杂度及观测数据的增加,需要越来越长的时间才能完成采样过程。MLE算法在不足1s内便能完成的运算而MCMC算法甚至需要几分钟时间才能完成,这对计算机内存和运算时间都是极大的考验。MCMC算法不能对生成的随机样本充分利用,在本文模型参数估计中的利用率仅为 $8/50=16\%$,造成了计算资源浪费。此外,采样算法收敛程度是衡量参数估计结果可靠性的重要指标,所以利用MCMC算法进行参数估计之前,需要对获得的随机样本链进行常规性收敛检验。最后,本文采用的切片采样算法在迭代过程中每个参数都条件依赖于其他参数而依次更新,这可能会增加随机样本的自相关程度并增加“Burn-in”与“Thinning”程序所需的样本量。如何改进算法(如将切片采样算法的参数更新方式由依次更新变为同步更新)从而加快算法收敛速度并提高资源利用率是今后研究的重点。

致谢:感谢Silke Schmidt博士、Helena Rodnight博士、欧先交博士、Amanda Keen-Zebert博士提供的用于模型验证的实测 De 数据。

参考文献 (References)

- 彭俊,董治宝,韩凤清,等. 2016. 基于R语言的光释光年代学数据处理程序包numOSL程序设计及应用实例分析[J]. 中国沙漠,待刊. [Peng J, Dong Z B, Han F Q, et al. 2016. Developing a package numOSL for analyzing optically stimulated luminescence data using the R statistical language and its practical application[J]. Journal of Desert Research, in press.]
- 彭俊,董治宝,张正偲. 2015. 基于Monte Carlo方法的释光

- 测年剂量率误差估计[J]. 核技术, 38(7): 070201. [Peng, J, Dong Z B, Zhang Z C. 2015. Determining the error of dose rate estimates on luminescence dating using Monte Carlo approach[J]. Nuclear Techniques, 38(7): 070201.]
- Arnold L J, Roberts R G. 2009. Stochastic modelling of multi-grain equivalent dose (De) distributions: Implications for OSL dating of sediment mixtures[J]. Quaternary Geochronology, 4(3): 204-230.
- Bateman M D, Boulter C H, Carr A S, et al. 2007. Detecting post-depositional sediment disturbance in sandy deposits using optical luminescence[J]. Quaternary geochronology, 2(1-4): 57-64.
- Duller G A T. 2007. Assessing the error on equivalent dose estimates derived from single aliquot regenerative dose measurements. Ancient TL, 25(1): 15-24.
- Duller G A T. 2008. Single-grain optical dating of Quaternary sediments: Why aliquot size matters in luminescence dating[J]. Boreas, 37(4): 589-612.
- Galbraith R F. 2003. A simple homogeneity test for estimates of dose obtained using OSL[J]. Ancient TL, 21(2): 75-77.
- Galbraith R F, Green P F. 1990. Estimating the component ages in a finite mixture[J]. International Journal of Radiation Applications and Instrumentation. Part D. Nuclear Tracks and Radiation Measurements, 17(3): 197-206.
- Galbraith R F, Roberts R G. 2012. Statistical aspects of equivalent dose and error calculation and display in OSL dating: An overview and some recommendations[J]. Quaternary Geochronology, 11: 1-27.
- Galbraith R F, Roberts R G, Laslett G M, et al. 1999. Optical dating of single and multiple grains of quartz from Jinmium rock shelter, northern Australia: Part I, experimental design and statistical models[J]. Archaeometry, 41(2): 339-364.
- Galbraith R F, Roberts R G, Yoshida H. 2005. Error variation in OSL palaeodose estimates from single aliquots of quartz: A factorial experiment[J]. Radiation Measurements, 39(3): 289-307.
- Gelman A, Carlin J B, Stern H S, et al. 2013. Bayesian data analysis[M]. 3rd ed. London, Britain: Chapman and Hall Press.
- Huntriss A. 2008. A Bayesian analysis of luminescence dating [D]. Durham, British: Durham University.
- Huntriss A. 2008. A Bayesian analysis of luminescence dating [D]. Durham University.
- Keen-Zebert A, Tooth S, Rodnight H, et al. 2013. Late Quaternary floodplain reworking and the preservation of alluvial sedimentary archives in unconfined and confined river valleys in the eastern interior of South Africa[J]. Geomorphology, 185: 54-66.
- Kunz A, Pflanz D, Weniger T, et al. 2013. Optically stimulated luminescence dating of young fluvial deposits of the Middle Elbe River Flood Plains using different age models [J]. Geochronometria, 41(1): 36-56.
- Lepper K, McKeever S W S. 2002. An objective methodology for dose distribution analysis[J]. Radiation Protection Dosimetry, 101(1-4): 349-352.
- Lunn D, Jackson C, Best N, et al. 2013. The BUGS book: A practical introduction to bayesian analysis[M]. London, Britain: Chapman and Hall press.
- Murray A S, Wintle A G. 2000. Luminescence dating of quartz using an improved single-aliquot regenerative-dose protocol[J]. Radiation Measurements, 32(1): 57-73.
- Neal R M. 2003. Slice sampling[J]. The Annals of Statistics, 31(3): 705-767.
- Neale M C, Miller M B. 1997. The use of likelihood-based confidence intervals in genetic models[J]. Behavior Genetics, 27(2): 113-120.
- Olley J, Caitcheon G, Murray A. 1998. The distribution of apparent dose as determined by optically stimulated luminescence in small aliquots of fluvial quartz: Implications for dating young sediments[J]. Quaternary Science Reviews, 17(11): 1033-1040.
- Ou X J, Duller G A T, Roberts H M, et al. 2015. Single grain optically stimulated luminescence dating of glacial sediments from the Baiyu Valley, southeastern Tibet[J]. Quaternary Geochronology, 30: 314-319.
- Peng J, Dong Z B, Han F Q, et al. 2013. R package numOSL: Numeric routines for optically stimulated luminescence dating[J]. Ancient TL, 31(2): 41-48.
- Peng J, Dong Z B. 2014. A simple Bayesian method for assessing the standard error of equivalent dose estimates[J]. Ancient TL, 32(2): 17-23.
- Peng J, Dong Z B, Han F Q, et al. 2014. Estimating the number of components in an OSL decay curve using the Bayesian Information Criterion[J]. Geochronometria, 41(4): 334-341.
- Pietsch T J. 2009. Optically stimulated luminescence dating of young (<500 years old) sediments: Testing estimates of burial dose[J]. Quaternary Geochronology, 4(5): 406-422.
- Plummer M, Best N, Cowles K, et al. 2006. Coda: Convergence diagnosis and output analysis for MCMC[J]. R News, 6(1): 7-11.
- Rhodes E J, Ramsey C B, Outram Z, et al. 2003. Bayesian methods applied to the interpretation of multiple OSL dates: High precision sediment ages from Old Scatness Broch excavations, Shetland Isles[J]. Quaternary Science Reviews, 22(10-13): 1231-1244.
- Roberts R G, Galbraith R F, Yoshida H, et al. 2000. Distin-

- guishing dose populations in sediment mixtures: A test of single-grain optical dating procedures using mixtures of laboratory-dosed quartz[J]. *Radiation Measurements*, 32(5-6): 459-465.
- Rodnight H, Duller G A T, Wintle A G, et al. 2006. Assessing the reproducibility and accuracy of optical dating of fluvial deposits[J]. *Quaternary Geochronology*, 1(2): 109-120.
- Schmidt S, Tsukamoto S, Salomon E, et al. 2012. Optical dating of alluvial deposits at the orogenic front of the Andean Precordillera (Mendoza, Argentina)[J]. *Geochronometria*, 39(1): 62-75.
- Schwarz G. 1978. Estimating the dimension of a model[J]. *The Annals of Statistics*, 6(2): 461-464.
- Sivia D S, Burbidge C, Roberts R G, et al. 2004. A Bayesian approach to the evaluation of equivalent doses in sediment mixtures for luminescence dating[J]. *AIP Conference Proceedings*, 735: 305-311.
- Sivia D S, Burbidge C, Roberts R G, et al. 2004. A Bayesian approach to the evaluation of equivalent doses in sediment mixtures for luminescence dating[J]. *Bayesian Inference and Maximum Entropy Methods in Science and Engineering*, 735: 305-311.
- Thomsen K J, Murray A S, Bøtter-Jensen L, et al. 2007. Determination of burial dose in incompletely bleached fluvial samples using single grains of quartz[J]. *Radiation measurements*, 42(3): 370-379.
- Van Der Touw J W, Galbraith R F, Laslett G M. 1997. A logistic truncated normal mixture model for overdispersed binomial data[J]. *Journal of Statistical Computation and Simulation*, 59(4): 349-373.
- Zhang J F, Zhou L P, Yue S Y. 2003. Dating fluvial sediments by optically stimulated luminescence: Selection of equivalent doses for age calculation[J]. *Quaternary Science Reviews*, 22(10-13): 1123 - 1129.
- Zink A. 2013. A coarse Bayesian approach to evaluate luminescence ages[J]. *Geochronometria*, 40(2): 90-100.

Application of slice sampling method for optimizing OSL age models used for equivalent dose determination

PENG Jun¹, DONG Zhibao¹, HAN Fengqing²

(1. Cold and Arid Regions Environmental and Engineering Research Institute, CAS, Lanzhou 730000, China;

2. Qinghai Institute of Salt Lakes, CAS, Xining 810008, China)

Abstract: In Optically Stimulated Luminescence (OSL) dating, statistical age models used for equivalent dose (De) determination are probabilistic models constructed according to mathematical statistics. They are applied to distinguish De populations that are sedimentologically different or to determine a De value that represents the burial dose of a sample. Maximum likelihood estimation (MLE) method is routinely used to optimize parameters of an age model. In the present study, we used the Slice sampling algorithm to determine the parameters of age models. Slice sampling is a Markov chain Monte Carlo (MCMC) sampling method, which enables the sampling distributions of parameters to be obtained from the joint likelihood function that is determined by observations and the specified model. This study applied easily implemented and openly accessible numeric routines to performing the algorithm. We used artificial and measured datasets to check the reliability of the estimates. MCMC method is insensitive to the parameters' initial states, and the standard errors (or confidence intervals) of parameters assessed using this method are more reliable compared to those based on the Fisher information matrix constructed through numerical differentiation. Our results indicate that the Slice sampling method provides an alternative for age model optimization. Slice sampling method generates an informative estimation for the results of MLE method in age model application.

Key words: OSL dating; equivalent dose; statistical age models; parameter optimization; stochastic sampling