Bayesian Chronological Models implemented to the R Package BayLum

Anne Philippe, Guillaume Guerin and Claire Christophe

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

Introduction : OSL dating

Problem Construction of a statistical model to estimate the age of a sample from OSL signals.

- Age equation:

$$\text{Age} = \frac{\text{palaeodose}}{\text{dose rate}}$$

For fixed sample i

- The parameter of interest is the age of the sample i denoted A_i
- The palaeodose of the sample i (denoted $\mathcal{D}_i$) is not observed
  $\rightsquigarrow$ estimation step is requided
- $\dot{d}_i \pm \dot{\sigma}_i$ dose rate of the sample and its estimation error.
  Errors on dose rate come from
  - measurement error
  - calibration error

Classical approach for an individual Age

1. We estimate the palaeodose of the sample using a Bayesian or frequentist method $\rightsquigarrow \widehat{\mathcal{D}}_i \pm \widehat{\sigma}_{\mathcal{D}_i}$
2. the dose rate is known with an uncertainty $\dot{d}_i \pm \dot{\sigma}_i$
3. We approximate the age by

$$\widehat{A}_i = \frac{\widehat{\mathcal{D}}_i}{\dot{d}_i} \quad \text{with error} \quad \sigma_{A_i} = \widehat{A}_i \sqrt{\left(\frac{\widehat{\sigma}_{\mathcal{D}_i}}{\widehat{\mathcal{D}}_i}\right)^2 + \left(\frac{\dot{\sigma}_i}{\dot{d}_i}\right)^2}$$

An Alternative approach is based on the Monte Carlo simulations [See Huntriss et al]
We sample the distribution of age A_i under the assumption that both variables D_i and $\dot{d}_i$ are Gaussian.

4. $\widehat{A}_i \pm \sigma_{A_i}$ can be included in chronological modelling. (e.g. models implemented to Oxcal ; Chronomodel ...)

Motivations for Bayesian modelling

- Joint estimation of a sequence of ages
- Take into account the systematic and the individual errors in the dose rate term.
- Include additional information (statigraphy) : temporal ordering between ages
- Combine different dating methods : add 14C age.

Solution implemented in BayLum a multivariate model with multiplicative Gaussian errors in a Bayesian framework.

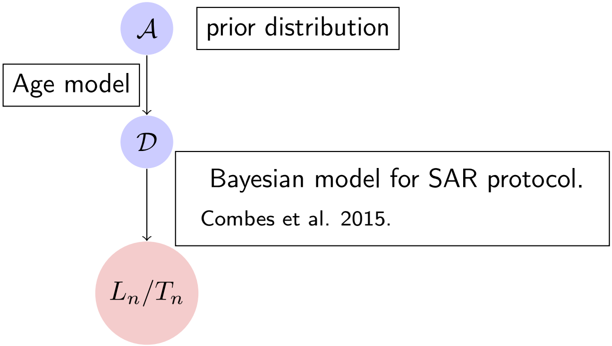
Reference :

`Benoit Combes and Anne Philippe`
`In Quaternary Geochronology, Volume 39, 2017, Pages 24-34`

Bayesian modelling

- Observations : OSL signal of each grain of each sample L_n/T_n
- Parameters of interest : Age $\mathcal{A} = (A^{(i)})_i$ of each sample
- Additional variables : the Palaeodose $\mathcal{D} = (\mathcal{D}_i)_i$ of each sample

The hierarchical structure is



Probability distribution is of the form :

$$P(\mathcal{A}, L_n/T_n, \mathcal{D}) = P(L_n/T_n|\mathcal{D})P(\mathcal{D}|\mathcal{A})P(\mathcal{A})$$

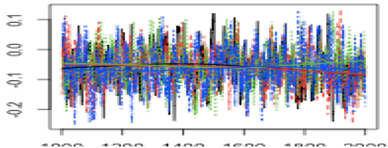
Bayesian modelling

1. Using Bayes Formula, we get the posterior density of $\mathcal{A}$:

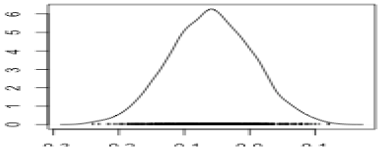
$$p(A|L_n/T_n) \propto p(A) \int P(L_n/T_n|D)P(D|A)\mathrm{d}\ D$$

2. No explicit form of the posterior distribution
3. MCMC algorithms are required to evaluate the posterior distribution.

Simulation of Markov Chains



Estimation of the posterior density



Age model

The statistical model is

$$\mathcal{D}_i = A_i(\dot{d}_i + \varepsilon_{\dot{d}_i} + \alpha_i \varepsilon_c)$$

where

- $\varepsilon_{\dot{d}_i} \sim \mathcal{N}(0, \sigma_{\dot{d}_i})$ =individual error (depends on the sample)
- $\varepsilon_c \sim \mathcal{N}(0, \sigma_c)$ = systematic error (common for all samples)
- $\alpha_i > 0$: weight of systematic error for sample i.

$\rightsquigarrow$ correlations between the palaeodoses $\mathcal{D}$

Conditionnally to the ages $\mathcal{A}$

$$(\mathcal{D}_i)_i \sim \mathcal{N}\left((A_i \dot{d}_i)_i, \Sigma\right)$$

where the covariance matrix Σ satisfies:

$$\Sigma_{i,i} = {A_i}^2 \left({\sigma_{\dot{d}_i}}^2 + {\alpha_i}^2 {\sigma_c}^2\right) \qquad \Sigma_{i,j} = A_i A_j \alpha_i \alpha_j {\sigma_c}^2$$

Prior distribution on the ages $\mathcal{A}$

We assume that all ages belong to the time interval $[a_{min}, a_{max}]$.
This is the study period.

1. Non informative prior on the ages [Jeffreys prior]

$$p(A_1, ..., A_N) = \prod_{i=1}^{N} p(A_i)$$
$$\propto \prod_{i=1}^{N} \frac{1}{A_i} \mathbb{1}_{[a_{min}, a_{max}]}(A_i)$$

2. temporal ordering on the sequence or subsequence of ages

$$A_1 \leq A_2 \leq ... \leq A_N$$

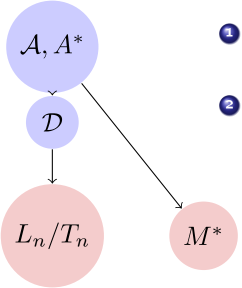
$$p(A_1, ..., A_N) \propto \prod_{i=1}^{N} \frac{1}{(A_i)} \mathbb{1}_{\mathcal{B}}(A_1, ..., A_n)$$

where $\mathcal{B} = \{A_1 \leq A_2 \leq ... \leq A_N\} \cap [a_{min}, a_{max}]^N$

Additionnal observations

We can integrate measurements coming from different chronometric techniques
Let M^* be the measurement and A^* the unknown age.

The model becomes



1. The measurements M^* and OSL signals are independent conditionnally to the ages.
2. The prior is given by

$$p(A_1, ..., A_N, A^*) \propto \prod_{i=1}^{N} \frac{1}{A_i} p(A^*) \, \mathbb{1}_{\tilde{\mathcal{B}}}(A_1, ..., A_n, A^*)$$

where $\tilde{\mathcal{B}}$ defines the temporal ordering on $\mathcal{A}, A^*$

Example : Radiocarbon dating.

The distribution of M^* conditionnally to A^* is

$$\mathcal{N}(g(A^*), s^2 + \sigma_g^2(A^*))$$

where

- g is the calibration curve and σ_g^2 the calibration curve error
- s^2 is the laboratory error

Outlier model can be added for instance using a mixture distribution

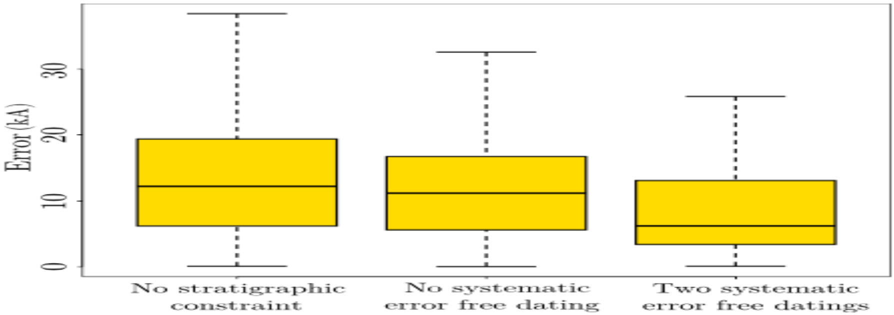
$$(1-p)\mathcal{N}(g(A^*), s^2 + \sigma_g^2(A^*)) + p\mathcal{N}(g(A^*), \alpha(s^2 + \sigma_g^2(A^*)))$$

- p is the probabiliy that M is an outlier.
- α is estimated if M is detected as an outlier.

Illustration and conclusion

We consider five samples and three settings:

- five samples without any stratigraphic constraints, all affected by the systematic error term
- the same five samples with stratigraphic constraints
- two of them are not affected by the systematic-error term (14C or OSL done in different labs)



Box plots of the estimation error the age 3 (estimation is based on 100 datasets)