Bayesian modelling applied to dating methods

Anne Philippe
Laboratoire de mathématiques Jean Leray
Université de Nantes, France
Anne.Philippe@univ-nantes.fr

Mars 2018, Muséum national d'Histoire naturelle

Master évolution, patrimoine naturel et sociétés
spécialité : quaternaire et préhistoire

Plan

Bayesian approach to Interpreting Archaeological Data

The statistical modelling within the Bayesian framework is widely used by archaeologists :

- 1988 Naylor , J . C. and Smith, A. F. M.
- 1990 Buck C.E.
- 1994 Christen, J. A.
- etc

Examples

- Bayesian interpretation of 14C results , calibration of radiocarbon results.
- Constructing a calibration curve.
  to convert a measurement into calendar date
- Bayesian models for relative archaeological chronology building.

Observations

Each dating method provides a measurement M, which may represent :

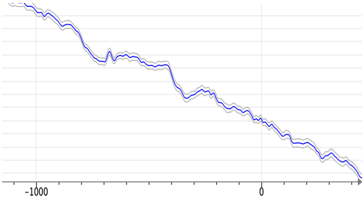
- a 14C age,
- a paleodose measurement in TL/OSL,
- an inclination, a declination or an intensity of the geomagnetic field

Relation with calendar date

$$M = g(\theta) + \epsilon$$

where

- θ is the calendar time
- g is a calibration function which relates the measurement to θ



Radiocarbon *IntCal14*

Archaeological information

After the archaeological excavations, prior information is available on the dates.

Examples :

- Dated archaeological artefacts are contemporary
- Stratigraphic Information which induces an order on the dates.
- the differences between two dates is known (possibly with an uncertainty).
- *Terminus Post Quem/ Terminus Ante Quem*
- etc

Bayesian statistics

- Observations $M_1, M_2, ...M_N$ whose the distribution depends on unknown parameter $f(M_1, ...M_n|\theta)$
- θ is the unknown parameters. We build a prior distribution on θ : $\pi(\theta)$

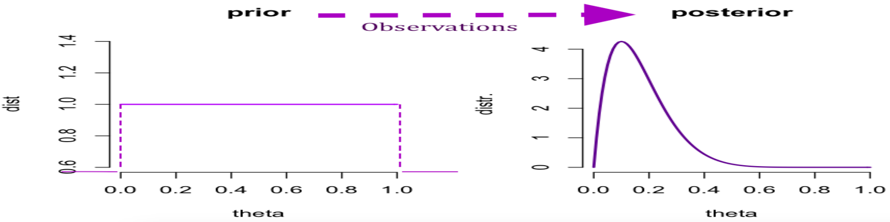
Example

- M_i : 14C ages done on artefact.
- θ : calendar date of artefact

Bayes Formula

The posterior distribution :

$$\pi(\theta|M_1, ...M_n) \propto f(M_1, ...M_n|\theta) \times \pi(\theta)$$



Example : Gaussian sample

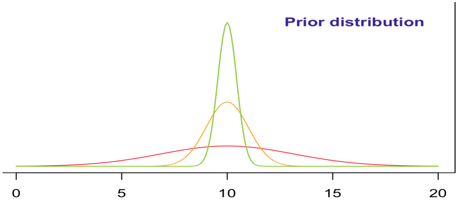
- Observation : n measurements with gaussian errors of the unknow quantity θ.

$$M_i = \theta + \epsilon \overset{iid}{\sim} \mathcal{N}\left(\theta, s^2\right)$$

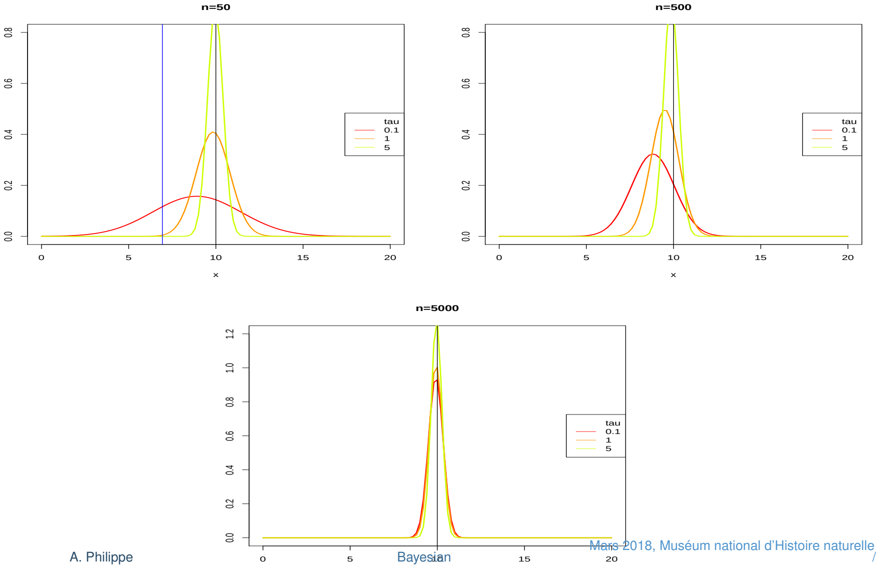
- Prior information on the unknown parameter θ : θ is close to 10 We translate this information as follows :

$$\theta \sim \mathcal{N}(10, \frac{1}{\tau})$$

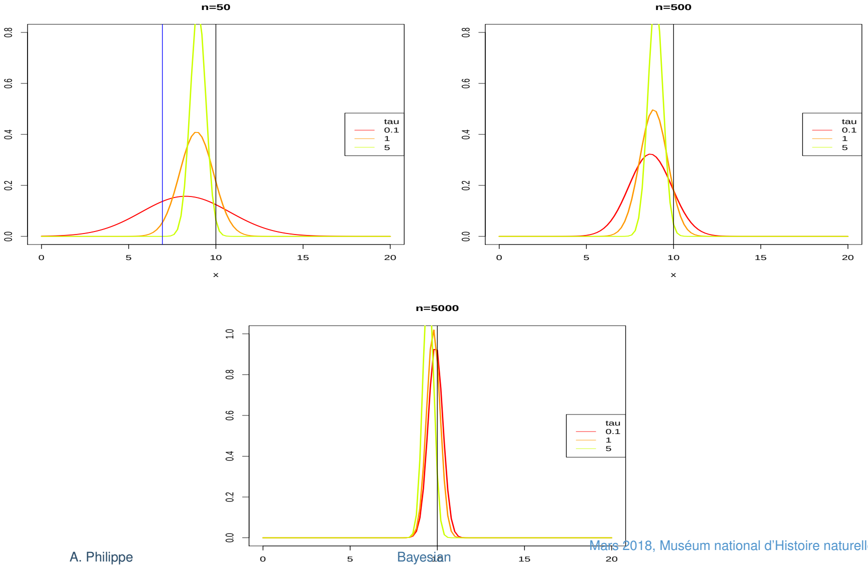
choice of τ ? it depends on our level of trust in the prior information



Behavior of he posterior distribution when the prior information is correct



Behavior of he posterior distribution : case of the wrong prior information

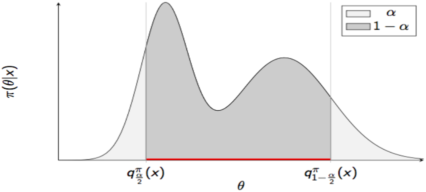


Bayesian inference

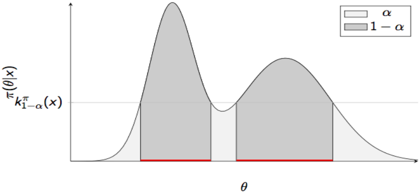
From the posterior distribution, we calculate

- Confidence region :

Credible interval



HPD region



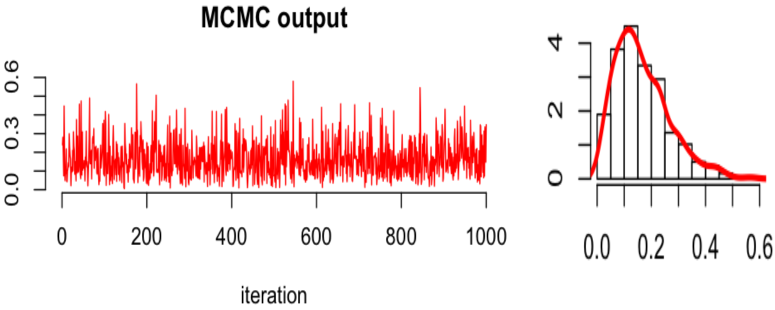
$$P(\theta \in \text{IC or HPD } |M_1, \ldots M_n) = 1 - \alpha$$

- Pointwise Estimates of the parameter *theta* :
  - Mean of the posterior distribution
  - Mode of the posterior distribution

Numerical approximation

Problem
An explicit form of the posterior distribution $\pi(\theta|M_1, ...M_n)$ is not available

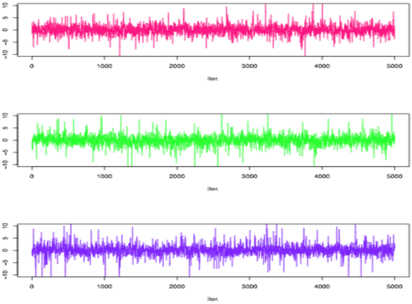
Solution
We simulate a sample using MCMC algorithm from the posterior distribution

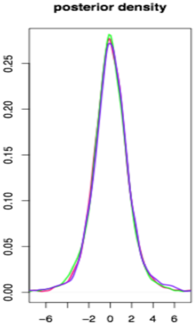


Diagnostic for MCMC output.

- History plots,
- marginal densities

All three graphical outputs act as diagnostic criteria to indicate a proper adaption of the model.





Softwares

1. `BCal` is an on-line Bayesian radiocarbon calibration tool.

   Buck C.E., Christen J.A. and James G.N. (1999). BCal an online Bayesian radiocarbon calibration tool. Internet Archaeology, 7

2. `Oxcal` provides radiocarbon calibration and analysis of archaeological and environmental chronological information.

   Bronk Ramsey, C. (1995). Radiocarbon calibration and analysis of stratigraphy The OxCal program. Radiocarbon, 37(2), 425-430.

3. `Chronomodel`

   Lanos, A. Philippe (2017) Hierarchical Bayesian modeling for combining Dates in archaeological context. Journal de la SFdS, Vol. 158 (2) pp 72-88.

   Lanos and Philippe (> 2018) Event date model a robust Bayesian tool for chronology building. Communications for Statistical Applications and Methods

R software

1. `ArchaeoPhases` Post-Processing of the Markov Chain Simulated by 'ChronoModel', 'Oxcal' or 'BCal'

   A. Philippe, M.-A. Vibet. (2017) Analysis of archaeological phases using the CRAN package ArchaeoPhases

2. `BayLum`. Chronological Bayesian Models Integrating Optically Stimulated Luminescence and Radiocarbon Age Dating

   B. Combes, A. Philippe. Bayesian analysis of individual and systematic multiplicative errors for estimating ages with stratigraphic constraints in optically stimulated luminescence dating. Quaternary Geochronology 39, 2017.

   A. Philippe, G. Guerin S. Kreutzer, BayLum an R package for Bayesian Analysis of OSL Ages & Chronological Modelling (LED2017)

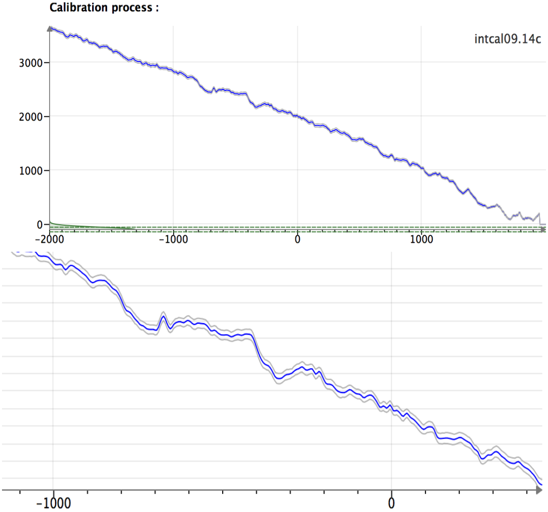
3. `ArchaeoChron` Bayesian Modeling of Archaeological Chronologies
4. `Luminescence` Comprehensive Luminescence Dating Data Analysis
5. `rbacon` age-modelling ; `Bchron` Radiocarbon Dating, Age-Depth Modelling

Plan

Different calibration curves

1. In radiocarbon :

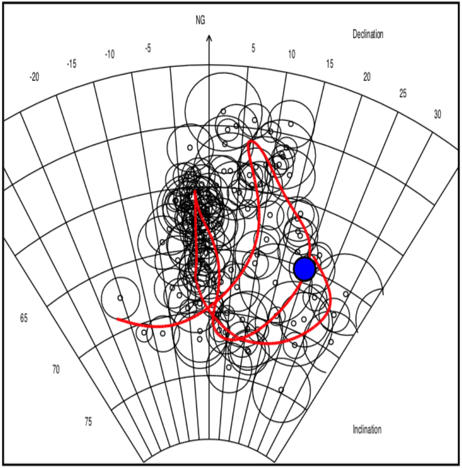
the curve *IntCal14* is used to convert an age measurement into calendar date for continental origin samples.



Different calibration curves

1. In radiocarbon :
2. In archaeomagnetism (AM),

the curve of secular variation of the geomagnetic field established for a given region are used to convert a measurement of inclination, declination or intensity into calendar dates.



Individual calibration

1. We observe M (14C, AM, TL/OSL measurement)

$$M = m + \epsilon$$

   where ϵ is the error of measurement. We assume $\epsilon \sim \mathcal{N}(0, s^2)$ where s is known.

2. Calibration : convert m $\rightarrow$ calendar date θ, the parameter of interest

$$m = g(\theta) + \sigma_g(\theta)\epsilon'$$

   where both functions g and σ_g are supposed known
   and where ϵ' represent the error on the calibration curve

3. Prior distribution on the parameter θ : Uniform distribution on T the study period.

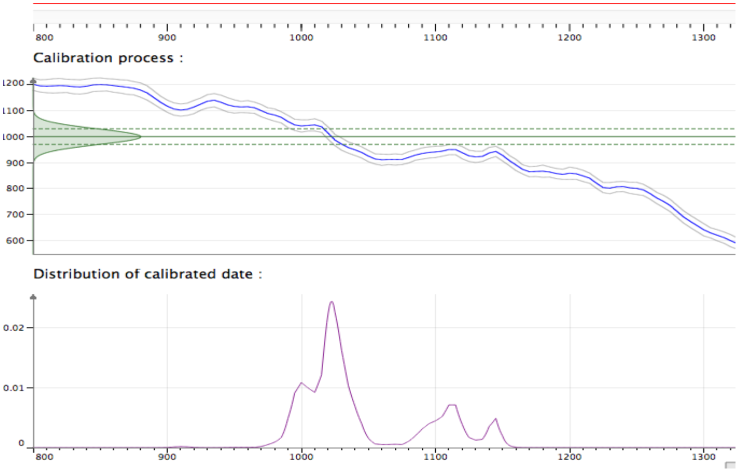
Posterior distribution :

$$p(\theta|M) \propto \frac{1}{S} \exp\left(\frac{-1}{2S^2} (M - g(\theta))^2 \right) 1_T(\theta)$$

where

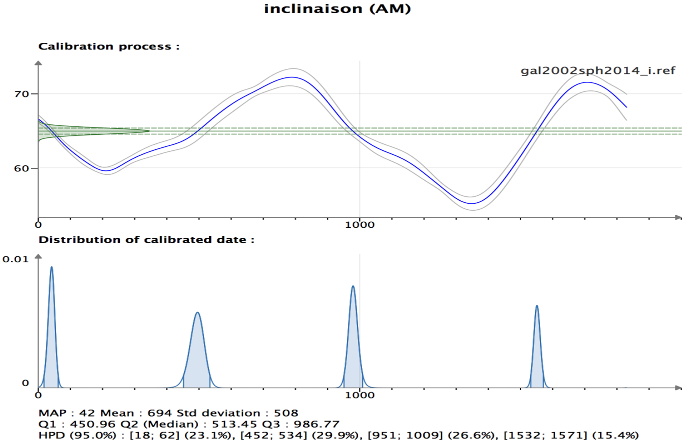
$$S^2 = s^2 + \sigma_g^2(\theta)$$

Radiocarbon



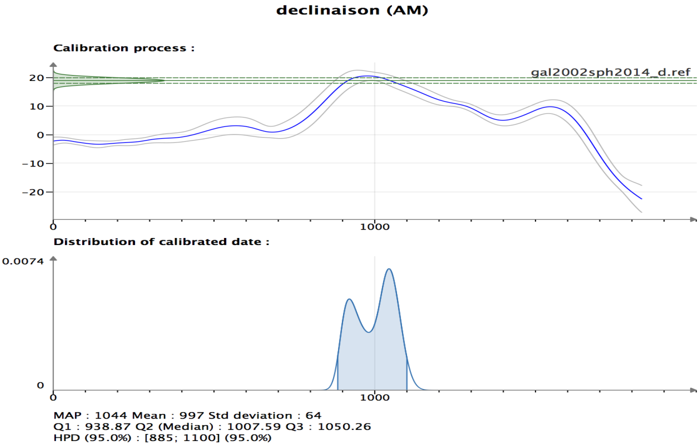
Converting a sample age 14C (= 1000 ± 30) in calendar date through the curve of Calibration *IntCal13*.

Archaeomagnetic calibration



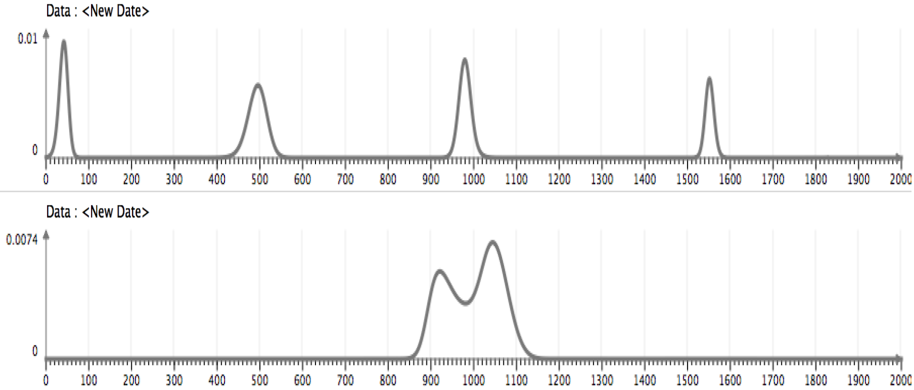
Converting an inclinaison measurement ($Incl = 65 \pm 1$) in calendar date via the calibration curve in France (Paris) over the last two millennia.

Archaeomagnetic calibration



Converting an declinaison measurement ($dec = 29$ with $Incl = 65 \pm 1$) in calendar date via the calibration curve in France (Paris) over the last two millennia.

Estimation of the date by two dating methods (Inclinaison / Declinaison)



How to combine the information coming from both dating methods to improve the accuracy of the estimated date ?

Plan

Definition of the target Event

Definition

- we choose a group of dated events that are related the target event.

$\rightsquigarrow$ Characterize the date of a target event from the combination of the dates of contemporaneous dates.

The objective is to estimate the calendar date of the "target event"
we denote θ the date of interest

The example of Lezoux

Medieval kiln of the potter's workshop in Lezoux (Auvergne, France)[1]

Aim : Dating the last firing of the kiln

[1] Menessier-Jouannet *et al.* 1995

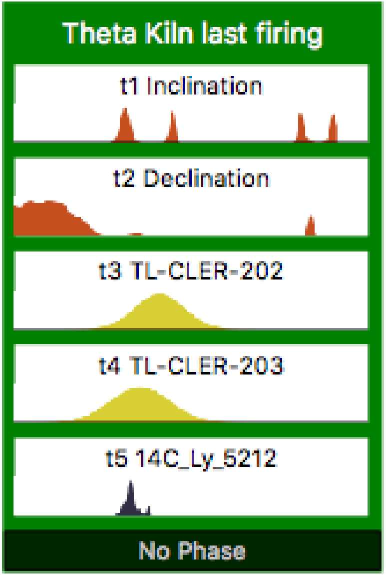
Lezoux - cont.

- **Target event** the date of the last firing (θ).

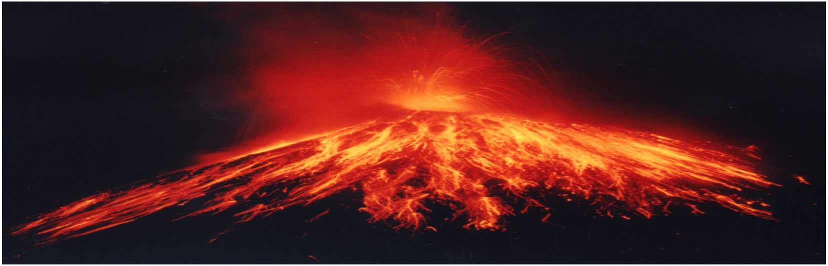
  This is any date between 0 and 2 000

- **dated events** :
  - baked clays dated by
    AM $>$ *Estimation of the last time the temperature exceeded a critical point*
    TL $>$ *Estimation of the last firing*
  - bones
    14C $>$ *Estimation of the death of the animal*

- **All these dated event are contemporaneous of the target event**



Volcanic eruptions



- Target Event : Eruptive period with flow deposits
- Dated events : organic samples found in a flow deposit are dated by 14C.

Definition of the Event Model

Lanos & Anne Philippe (2017,2018+)

1. We want to estimate θ. the date of the target event.

2. The target event is defined by
   - n measurements : $M_1, ..., M_n$
   - For each $i = 1, ..., n$ the measurement M_i is done on material whose calendar date t_i is unknown.

3. The prior information is

   the date of the target event belongs to $T = [T_b; T_e]$

$\rightsquigarrow$ we choose $T = [T_b; T_e]$ as study period .

The statistical model

The model is

$$
\begin{aligned}
M_i =& g_i(t_i) + \epsilon_i \\
t_i =& \theta + \lambda_i \\
& \theta \sim \text{Uniform } (T)
\end{aligned}
$$

Assumptions on ϵ_i :

ϵ_i represents the experimental and calibration error $\epsilon_i \sim_{ind} \mathcal{N}(0, s_i^2 + \sigma_g(t_i))$

Assumptions on λ_i :

λ_i represents the difference between the date of artifacts t_i and the target event θ This error is external to the laboratory.

$$\lambda_i \sim_{ind} \mathcal{N}(0, \sigma_i^2)$$

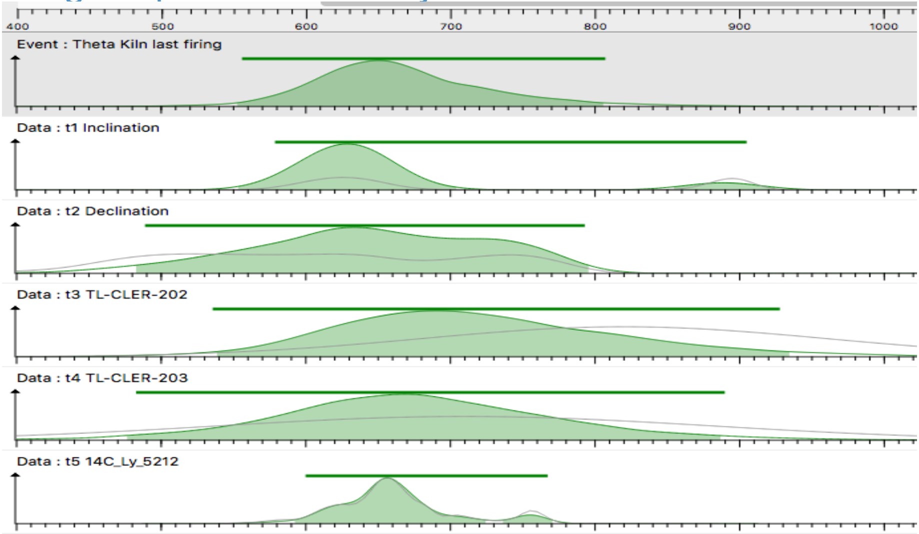
$\rightsquigarrow$ σ_i is the central parameter to ensure the robustness

Numerical result for Lezoux example.

- **Measurements**
  T1 : (AM) Inclination : I = 69.2, alpha = 1.2
  T2 : (AM) Declination : I = 69.2, alpha = 1.2, D= -2.8
  T3 : (TL) age 1170 +/- 140 years - Reference year : 1990
  T4 : (TL) age 1280 +/- 170 years - Reference year : 1990
  T5 : (14C) age 1370 +/- 50 BP

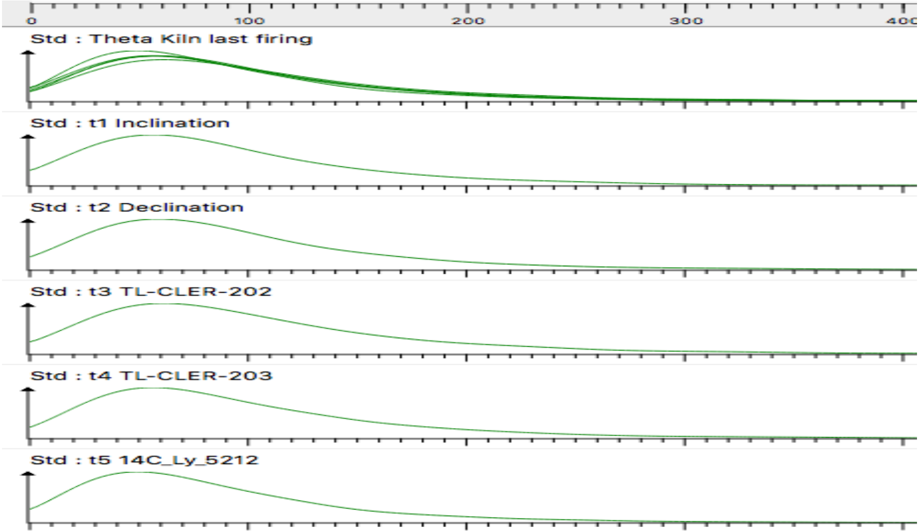
- **Prior information** We assume that the study period is [0 ; 2 000]

Marginal posterior density of the Event



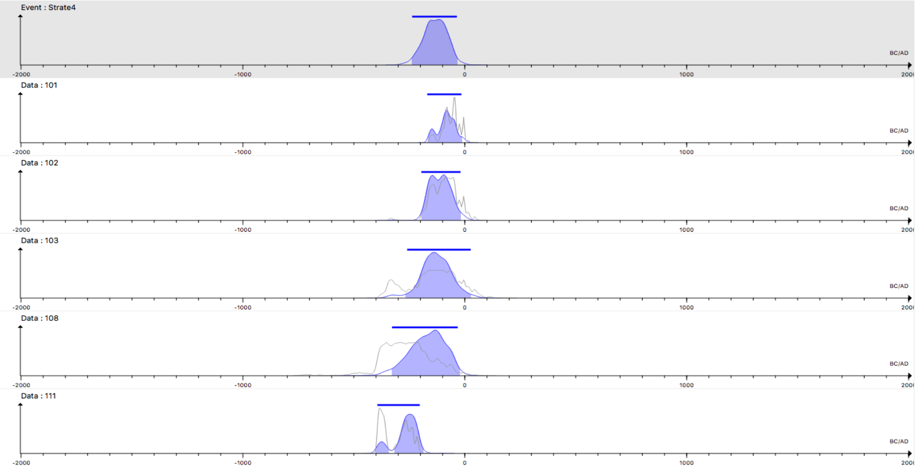
The segment above the curve represents the smallest credible interval.
The HPD region is presented by the colored area under the curve.

Marginal posterior densities of the individual standard deviations σ_i

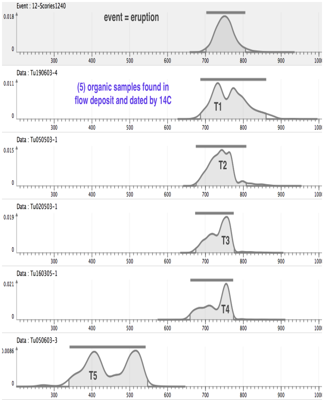


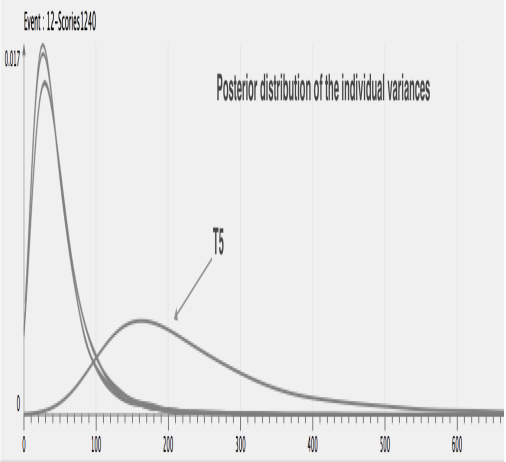
Numerical result for one pyroclastic flow

- Target event : eruption [θ]
- 5 organic samples found in flow deposit are dated by 14C [$t_1, ..., , t_5$]



Robustness of event model





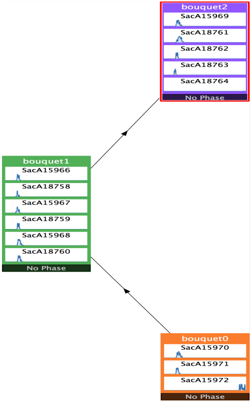
- the posterior density of date of the target Event remains almost insensitive to the outlier.
- We do not have to choose specific tools for rejecting outlying data.

Plan

We consider Bayesian tools for constructing chronological scenarios.

Main idea of the model implemented in `Chronemodel`

1. we define target event as a group of contemporaneous dated events.
2. We construct a chronology (= collection of dates) of target events taking into account temporal relationship between the dates of target events

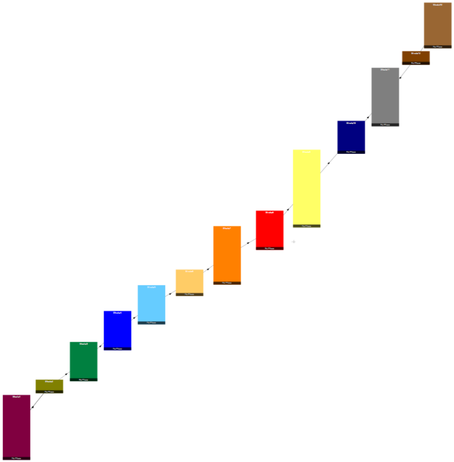


Alternative : model implemented in `Oxcal`

- We construct a chronology of dates of target events

Volcanic eruptions

- Target Event : Eruptive period with flow deposits
- Dated artefacts : organic samples found in a flow deposit are dated by 14C.
- Prior information Stratigraphic constraint on deposits



Restrictions

- Each event contains at least one measurrement.
- Each measurement is associated to one (and only one) target event.

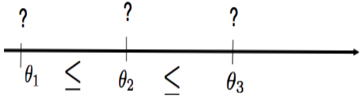
Chronologies of K target events

- We want to estimate $\theta_1, ...\theta_K$ the calendar dates of target events.

Prior information on the dates of the target event

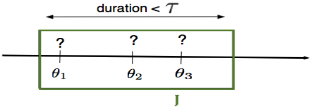
1. The stratigraphic constraints.

$\rightsquigarrow$ a partial order on $(\theta_1, ...\theta_K) := \vartheta \subset T^K$



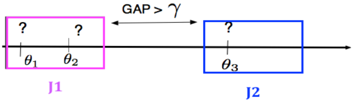
2. Duration information :

$\max_{j\in J} \theta_j - \min_{j\in J} \theta_j \leq \tau$ where τ is known



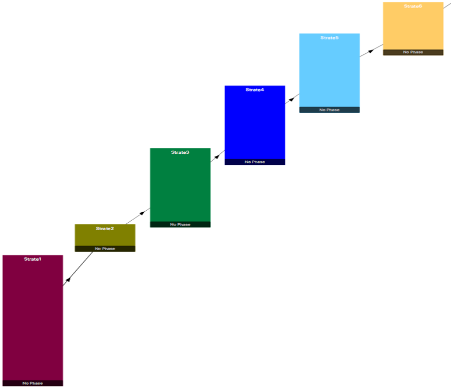
3. Hiatus information :

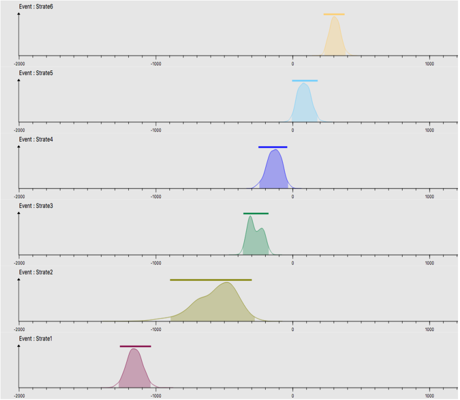
J_1, J_2 two groups, $\min_{j\in J_2} \theta_j - \max_{j\in J_1} \theta_j \geq \gamma$ where γ is known



Chronology of Volcanic eruptions

6 pyroclastic flows from volcano dated by 14C $\rightsquigarrow$ 6 ordered target events
$S = \{\vartheta : \ \theta_1 \leq \ldots \leq \theta_6\}$





Maya city with information on occupation time

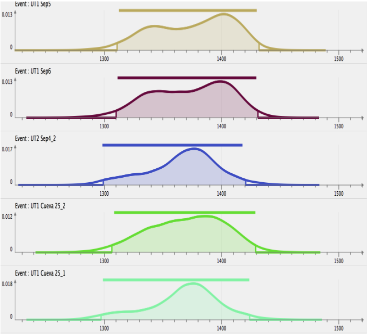
Prior information on the archaeological phase :

The occupation time is smaller than 50 years.

Comparison : HPD regions and posterior densites

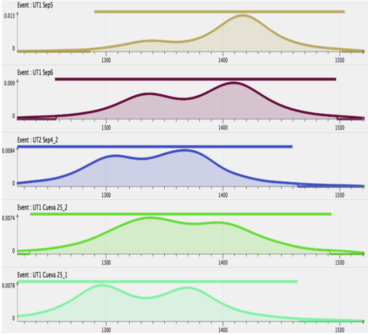
Prior information on the duration

θ_1	1309	1433
θ_2	1308	1430
θ_3	1299	1423
θ_4	1305	1429
θ_5	1297	1425



without prior information

θ_1	1284	1506
θ_2	1253	1502
θ_3	1213	1469
θ_4	1230	1497
θ_5	1192	1469



Plan

Description of the R package ArcheoPhase :

THis R package has its web interface

- Compatible with Oxcal or Chronomodel.
- The inputs are MCMC samples generated by both softwares.

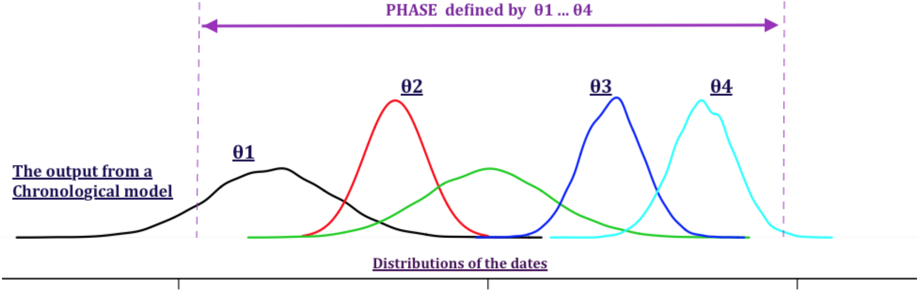
This package contains Statistical Tools for analysis the chronological modelling

Examples

1. Characterisation of a group of dates [begin / end /duration/ period]
2. Testing the presence of hiatus between two dates or two groups of dates.
3. Construction of tempo plot to evaluate the repartition in time

Phases : definition

A phase is a group of dates defined on the basis of objective criteria such as archaeological, geological or environmental criteria.



The collection of dates is estimated from a chronological model. [Chronomodel / Oxcal ...]

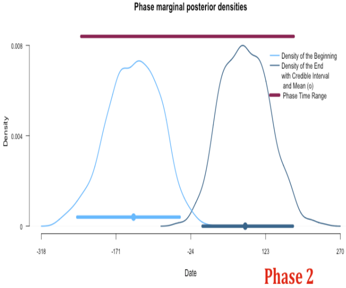
$$\text{Phase} = \{\theta_j,\ j \in J \subset \{1, ..., K\}\}$$

Estimation of the phase

Phase$_1 = \{\theta_j,\ j \in J \subset \{1, ..., K\}\}$.

1. posterior distribution of the minimum
   $\alpha = \min_{j \in J} \theta_j$
   $\rightsquigarrow$ Estimation of the beginning

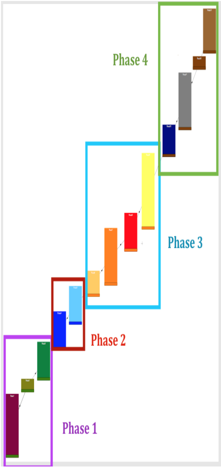
2. posterior distribution of maximum
   $\beta = \max_{j \in J} \theta_j \rightsquigarrow$ Estimation of the end

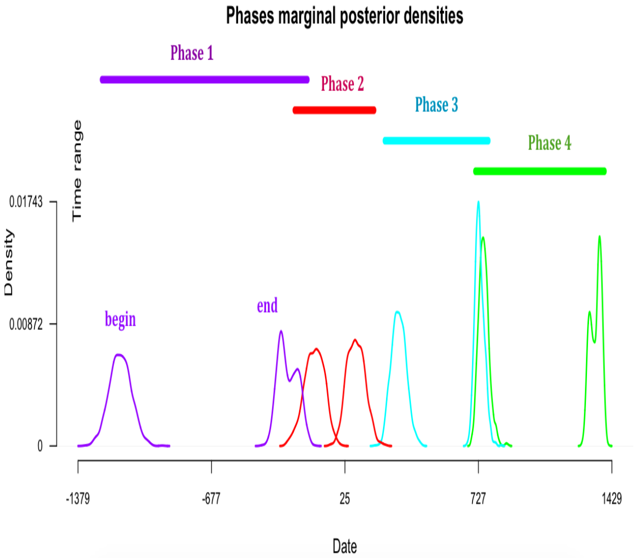


3. Phase time range The shortest interval that covers all the dates θ_j included in the phase at level 95%
   i.e. the shortest interval $[a, b] \subset T$ such that

$$P(\text{for all } j\ \theta_j \in [a, b] | M_1, ..., M_n) = P(a \leq \alpha \leq \beta \leq b | M_1, ..., M_n) = 95\%$$

Application to Volcanic eruptions [cont]

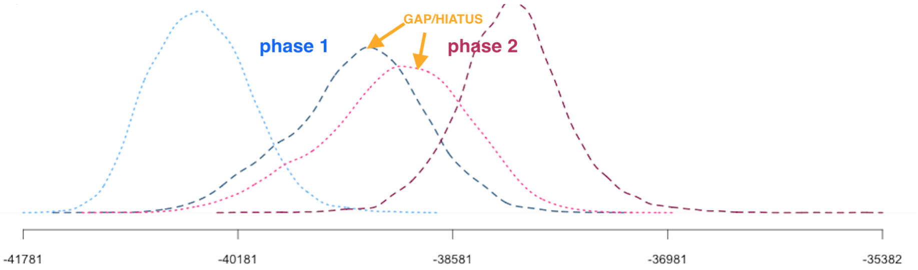




$$P_1 = \{\theta_1, \theta_2, \theta_3\}, \cdots\cdots$$

$$P_4 = \{\theta_{10}, \theta_{11}, \theta_{12}, \theta_{13}\}$$

Hiatus



Detection of a hiatus between two phases $\theta_j,\ j \in J_1$ and $\theta_j,\ j \in J_2$

1. $\beta_1 = \max_{j \in J_1} \theta_j$ and $\alpha_2 = \min_{j \in J_2} \theta_j$
2. Can we find $[c, d]$ such that

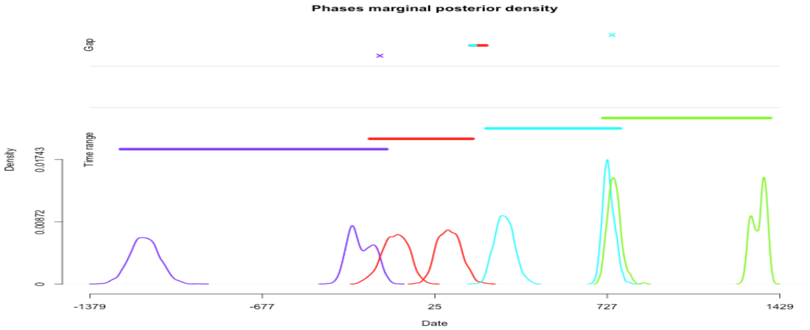
$$P(\beta_1 < c < d < \alpha_2 | M_1, ..., M_n) = 95\%?$$

Application cont.

Detection of hiatus :

- A hiatus is detected between Phases 2 & 3. Estimation of the interval [170, 235]
- there is no gap between 1 & 2 and 3 & 4

To summarise



The chronology of Canímar Abajo in Cuba

(Rocksandic *et al.* 2015 Philippe & Vibet (2018) RadioCarbon.

The site has evidence for two episodes of burial activity separated by a shell midden layer.

- 12 AMS radiocarbon dates (human bones collagen and a charcoal) obtained from burial contexts
- 7 from the Older Cemetery (OC),
- 5 from the Younger Cemetery (YC))

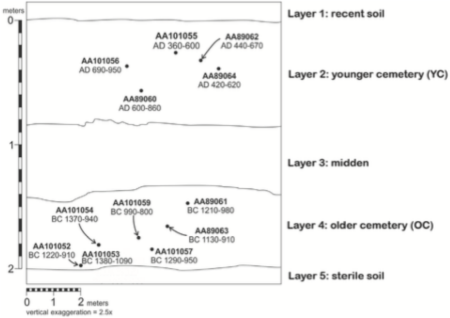
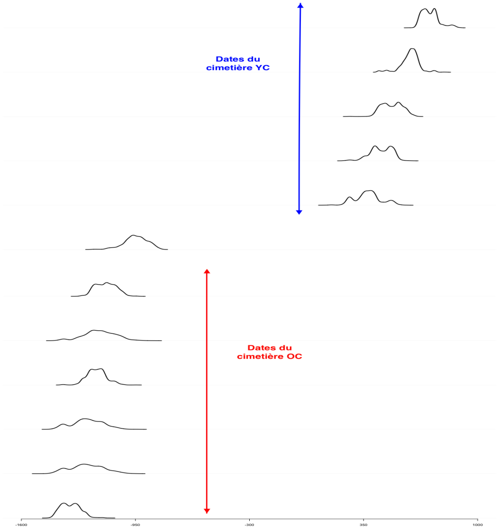


Figure 2 Stratigraphic profile indicating relative positions of samples for AMS [14]C dating

The aim : Bayesian model based on these 12 AMS radiocarbon dates in order to draw conclusions about

- the time of both mortuary activities
- the hiatus between them

The chronology

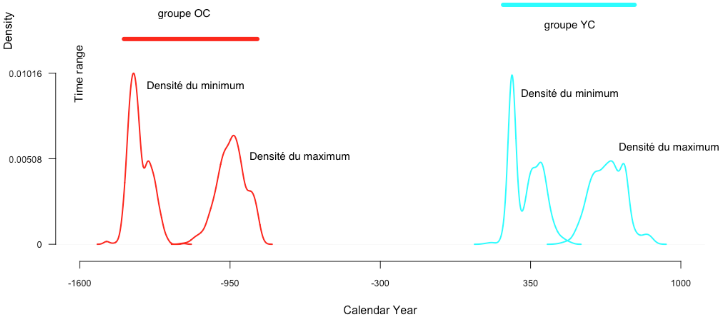


Estimation of the dates t

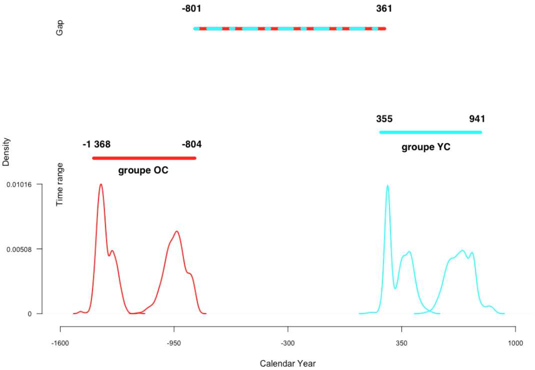
From the estimation of the sequence of dates $t_1, ..., t_{12}$ (using Bayesian model) we estimate

- the beginning and the end of the Older Cemetery
- the beginning and the end of the Younger Cemetery
- the gap between these two periods

Chronology of the activities in the site of Canimar Abajo.



Estimation of the gap



Testing the hypothesis "a date belongs to a time interval"

- We fix a time interval $[a, b]$.
- we want to test if the estimated date τ_1 belongs to this interval.

- In a Bayesian context, this consists in calculating the posterior probability :
$$P(a < \tau_1 < b | \mathcal{M})$$
- This probability gives the credibility of the hypothesis "the date τ_1 belongs to $[a, b]$".

Application.

We apply the testing procedure to allocate the 8 conventional radiocarbon dates to the most credible period among the five periods : before OC, OC, Midden period, YC and after YC.

Remark

We did not use these dates to construct the chronology of the site

Conventional radiocarbon dates	**Sampling level**	**Stratigraphic layer**	**Before OC**	**OC**	**Midden**	**YC**	**After YC**
UNAM.0714a	0.2 m	2 / YC	0	0	0	0	**100**
UNAM.0717	0.4 m	3 / midden	0	0	**100**	0	0
UNAM.0716	0.45 m	3 / midden	**100**	0	0	0	0
UNAM.0715	0.6-0.7 m	3 / midden	**100**	0	0	0	0
A.14315	0.9-1.0 m	3 / midden	0	0	**100**	0	0
UBAR.170	1.6-1.7 m	4 / OC	**100**	0	0	0	0
A.14316	1.8-1.9 m	4 / OC	0	**100**	0	0	0
UBAR.171	1.8-1.9 m	4 / OC	**100**	0	0	0	0

Sampling information and posterior probability for the the 8 conventional radiocarbon dates to belong to the periods of the chronology. Results are in %.

Tempo plot

(see Dye 2016 and Philippe & Vibet 2017)

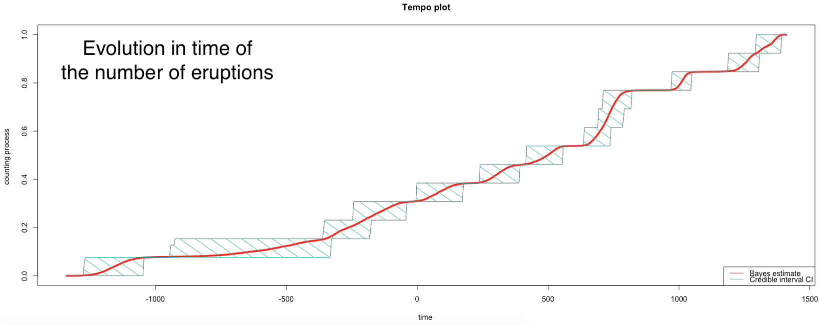
A statistical graphic designed for the study of rhythms.

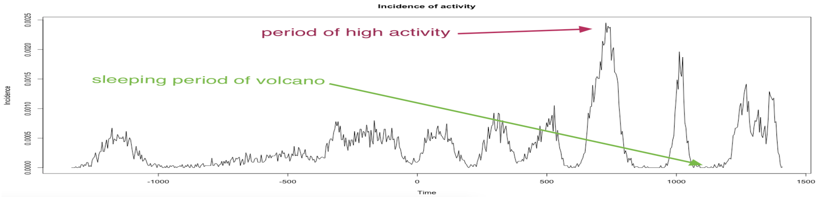
- The tempo plot measures change over time :
- For each date t, we estimate the number of events $N(t)$ which occurs before the date t, we have

$$N(t) = \sum_{i=1}^{n} \mathbb{I}_{]-\infty,t]}(\tau_i)$$

- Interpretation : the slope of the plot directly reflects the pace of change :
  - a period of rapid change yields a steep slope
  - a period of slow change yields a gentle slope.
  - When there is no change, the plot is horizontal.

Application : Evaluation of the activity of volcano





Age-depth model

Additinal information : the depth of the dated event.

1. We estimate the relation between the dates t and the depth h

$$f(t) = h \qquad \text{age-depth curve}$$

2. We estimate f taking into
   - all the posterior information on the sequence of dates estimated by the Bayesin chronological model
   - Non parametric regression method is applied on the output of the MCMC algorithm.
3. From the estimated curve, we predict the date as function of the depth.

Application : Age -depth curve and forecasting