L M J L
UMR 6629 - Nantes

Rencontres Mathématiques
de Rouen 2016

Statistique Bayesienne appliquée à la datation en Archéologie

Anne Philippe

Travail en collaboration avec

- Marie-Anne Vibet (Univ. Nantes)
- Philippe Lanos (CNRS, Bordeaux)

Plan

Bayesian analysis to Interpreting Archaeological Data

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

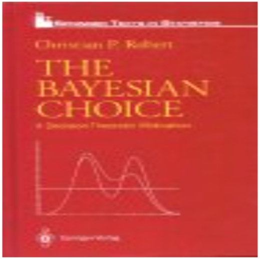
- 1988 Naylor , J . C. and Smith, A. F. M.
- 1990 Buck C.E.
- 1994 Christen, J. A.
- etc

Examples

- Bayesian interpretation of 14C results , calibration of radiocarbon results :
- Constructing a calibration curve.
  to convert a measurement into calendar date
- Constructing a Age-Depth relationship
  to relate the sediment depth with its age.
- Bayesian models for relative archaeological chronology building.

Bayesian approach

The Bayesian Choice
(1994,2001)
X'ian Robert



Parametric model : $X_1, ...X_n \sim P_\theta^{(n)}$ with $\theta \in \Theta$

Principle : θ is a random variable with probability distribution π.

⇝ uncertainty to the unknown parameter θ.

- Choice of π : the prior distribution π corresponds to our previous knowledge or belief about θ before observing the dataset

Bayesian Model :

1. Prior distribution : $\theta \sim \pi$
2. Conditionally on θ : $\mathbb{X} = (X_1, ...X_n) \sim P_\theta^{(n)} = P^{(n)}(\cdot|\theta)$

⇝ The posterior distribution of θ

$$\pi(\theta|x_1, ..., x_n) \propto p^{(n)}(x_1, ..., x_n|\theta)\pi(\theta)$$

Bayesian inference

Bayes estimate :

Under weak assumptions of regularity on the model :

- If θ_0 is the true parameter, then we have
  - $E(\theta|\mathbb{X}) \to \theta_0$ almost surely.
  - rate of convergence : $\sqrt{n}(E(\theta|\mathbb{X}) - \theta_0) \to N(0, I^{-1}(\theta_0))$ (in distribution)

Confidence region :

- Fix $\alpha \in (0,1)$. Confidence region I such that

$$P(\theta \in I|X_1, ..., X_n) = 1 - \alpha.$$

where I is

  - the smallest interval
  - or the highest probability density (HPD) region

$$I = \{\theta|\pi(\theta|x, ..., x_n) > C(\alpha)\}$$

Observations

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

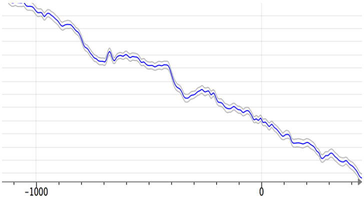
- a 14C age,
- a paleodose measurement in TL/OSL,
- an inclination, a declination or an intensity of the geomagnetic field

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

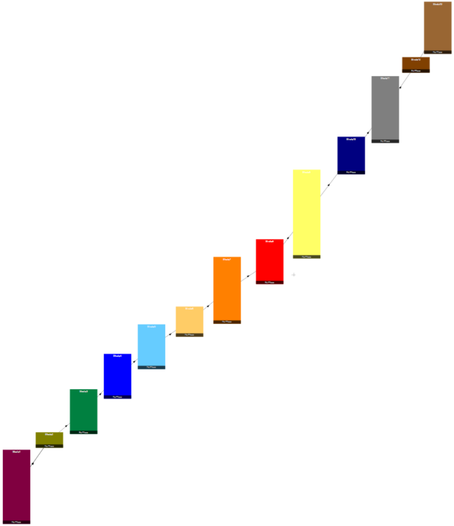
After the archaeological excavations, prior information is available on the dates.

Examples :

- Dated archaeological artefacts are contemporary, and so they define an archaeological event
- Stratigraphic Information which induces an order on the dates.
- the differences between two dates is known (possibly with an uncertainty).
- etc

Example : Volcano in Equador

- Event :
  Eruptive period with flow deposits down to the base of the volcano.
- Stratigraphic constraint on deposits



Plan

Medieval kiln of the potters in Lezoux (Auvergne)

1. Aim : Dating the last firing of the kiln
2. Prior information : any date between 0 and 2 000
3. Material found :
   - baked clays dated by
     - AM : Estimation of the last time the temperature exceeded a critical point
     - TL : Estimation of the last firing
   - a charcoal
     - 14C : Estimation of the felling of the tree

Event Model

Lanos & Philippe (2016)

- For each $i = 1, ..., n$: the measurement M_i is done on archaeological artefact with unknown calendar date t_i
- It is assumed that all archaeological artefacts are contemporary θ.

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

Assumptions on λ_i :

λ_i represents the error between t_i and θ due to sampling problems external to the laboratory

$$\lambda_i \sim_{ind} \mathcal{N}(0, \sigma_i^2) \qquad \rightsquigarrow \text{Individual effects}$$

Assumptions on ϵ_i :

ϵ_i represents the experimental error with variance provided by the labora tory s_i^2 and the calibration error

Prior on the individual variance : Non informative approach

- Jeffreys prior does not provide suitable solution.
- The prior distribution is improper and the posterior is not defined.
- Alternative : shrinkage Daniels (1999).

$$s_0^2/(\sigma_i^2 + s_0^2) \sim \mathcal{U}(0,1) \qquad s_0 \text{ is fixed}$$

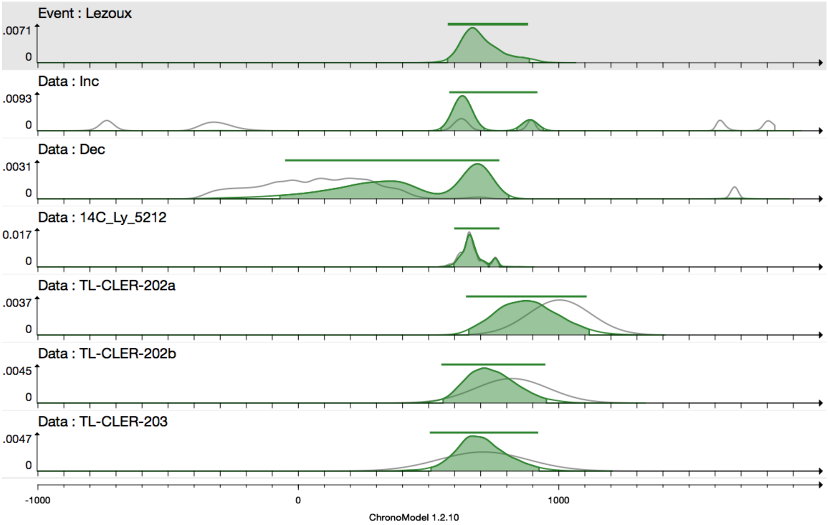
The density is of the form

$$p(\sigma_1^2, ..., \sigma_n^2) = \prod_{i=1}^{n} \frac{s_0^2}{(s_0^2 + \sigma_i^2)^2}.$$

Choice of s_0

- We fix s_0^2 as an estimate of the magnitude of error on the measurements.
- As s_0^2 is the median of the uniform shrinkage prior :
  this choice ensures that the parameter σ_i^2 has the same prior probability to be smaller or larger than s_0^2.

Illustration : Dating the last firing of the kiln Lezoux



Plan

Bayesian approach

Bayesian Central age model

Chronologies

Chronologies

- We want to build the chronology of K events
  $\rightsquigarrow$ a joint estimation of $\theta_1, ...\theta_K$ the calendar dates of the events.
- An event is defined by a collection of dated artifacts.

Prior information

1. The stratigraphic constraints. They imply a partial order on $(\theta_1, ...\theta_K) := \vartheta$
   $\rightsquigarrow S \subset T^K$
2. Duration information : Let $J \subset \{1, ..., K\}$

$$\max_{j\in J} \theta_j - \min_{j\in J} \theta_j \leq \tau$$

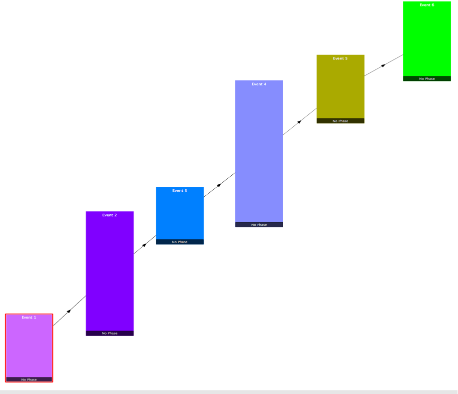
   τ is known
3. Hiatus information : Let $J_1, J_2 \subset \{1, ..., K\}$, $J_1 \cap J_2 = \emptyset$

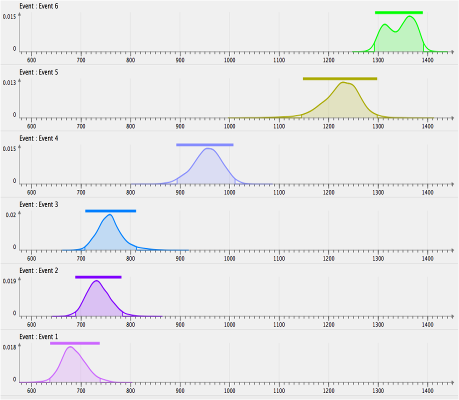
$$\min_{j\in J_2} \theta_j - \max_{j\in J_1} \theta_j \geq \gamma$$

volcano : pyroclastic flows

6 pyroclastic flows from volcano (Ecuador) dated by 14C $\rightsquigarrow$ 6 ordered events

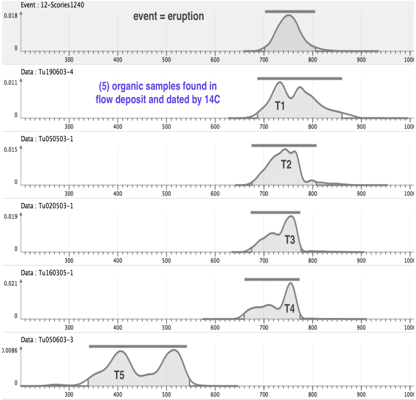
$S = \{\vartheta : \ \theta_1 \leq ... \leq \theta_6\}$

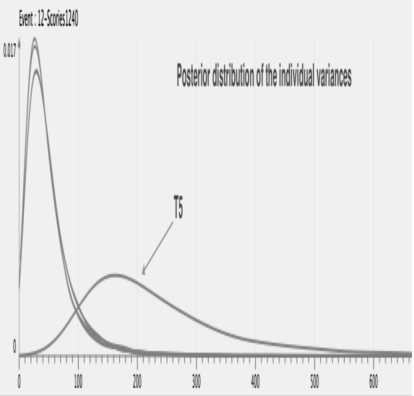




Robustness of event model

Focus on one pyroclastic flow





- Event posterior density remains almost insensitive to the outlier.
- Event model appears to be a robust statistics for calculating posterior mean of the date θ.

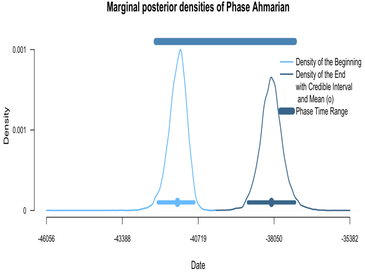
Phases

Definition

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

Estimation of the phase $\theta_j,\ j \in J \subset \{1, ..., K\}$

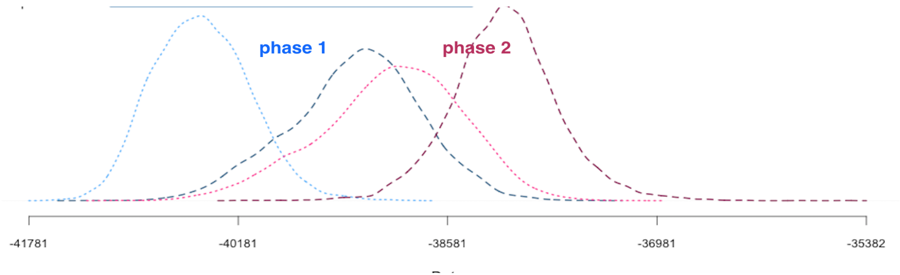
1. posterior distribution of $\alpha = \min_{j\in J} \theta_j \rightsquigarrow$ Estimation of the beginning
2. posterior distribution of $\beta = \max_{j\in J} \theta_j$ $\rightsquigarrow$ Estimation of the end



3. Phase time range The shortest interval that covers α and β at $100(1-\gamma)\%$ i.e. the shortest interval $[a, b] \subset T$ such that

$$P(a \le \alpha \le \beta \le b|M) = 1 - \gamma$$

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. Fix $\gamma (= 95\%)$. Can we find $[b, a]$ such that

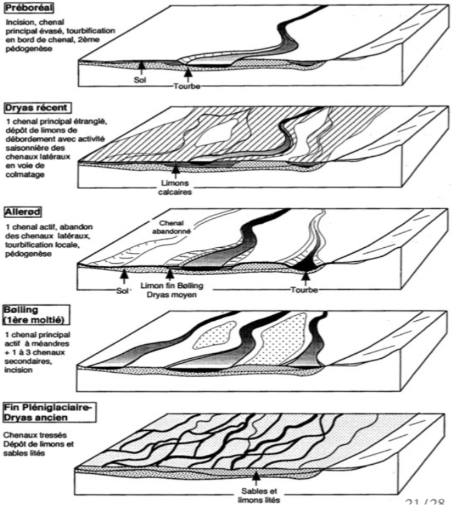
$$P(\beta_1 < b < a < \alpha_2 | \text{observations}) = \gamma?$$

Applications : Palynozones

Lateglacial pollen zones in the Paris basin[4]

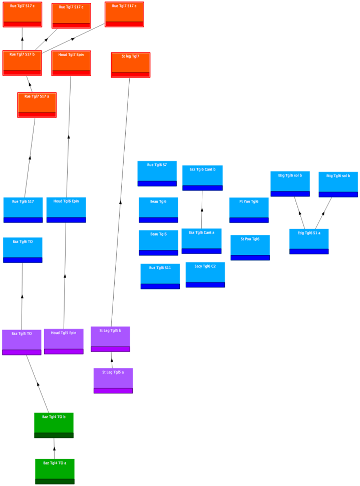
Aim : Defining chronological transitions between 4 phases

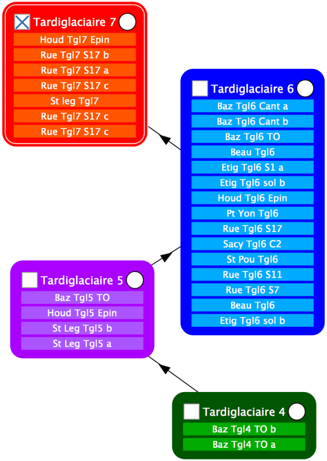
Tgl 7 : the younger Dryas
Tgl 6 : the second part of Allerød
Tgl 5 : the first part of Allerød
Tgl 4 : the older Dryas



[4] Leroyer *et al.* 2011, 2014

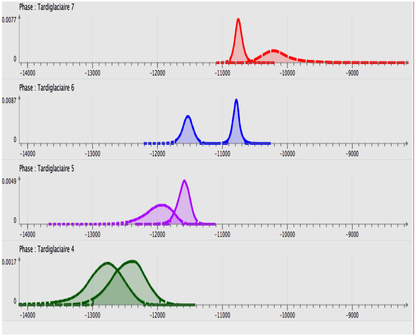
Tardiglaciare



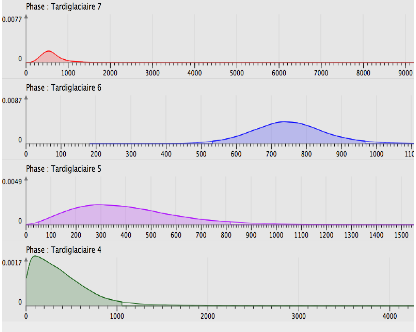


Estimation

Posterior distribution of α (beginning) and β (end)



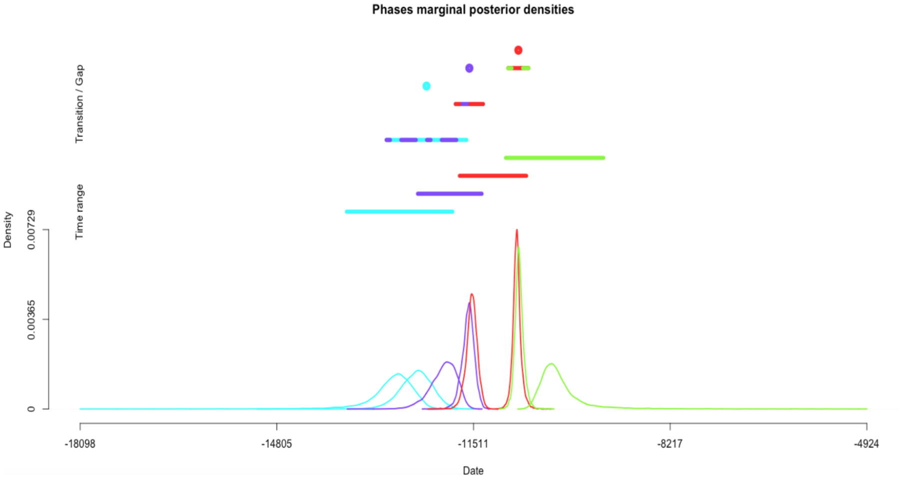
posterior distribution of duration

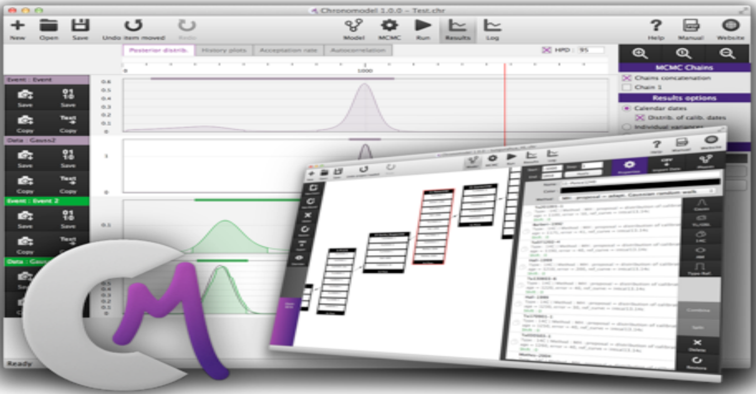


Detection of hiatus

there is no gap between two successive phases.

To summarise





www.chronomodel.fr

RChronoModel