ChronoModel

User manuel

Version : 1.1
Date : June 15, 2015

Director of project: Philippe Lanos (CNRS)
Co-director : Anne Philippe (Univ Nantes)

Authors, Affiliation, Email:
Marie-Anne Vibet, Univ.Nantes & CHL, marie-anne.vibet@univ-nantes.fr
Anne Philippe, Univ.Nantes, anne.philippe@univ-nantes.fr
Philippe Lanos, CNRS, philippe.lanos@univ-rennes1.fr
Philippe Dufresne, CNRS, philippe.dufresne@u-bordeaux-montaigne.fr

To cite Chronomodel in publications use:

Ph. Lanos, A. Philippe, H. Lanos, Ph Dufresne (2015). Chronomodel : Chronological Modelling of Archaeological Data using Bayesian Statistics. (Version 1.1). Available from http://www.chronomodel.fr.

A BibTeX entry for LaTeX users is

```
@Manual{chronomodel,
      title  = {Chronomodel  : Chronological Modelling of
               Archaeological Data using Bayesian Statistics.},
      author = {Ph. Lanos and  A. Philippe and H. Lanos
                and  Ph. Dufresne },
      year   = {2015},
      url    = {http://www.chronomodel.fr },
  }
```

Contents

Chapter 1

Introduction

Chronological modelling with « ChronoModel »

- Events:
  ChronoModel is based on the concept of “Event”. An Event is a point in time for which we can define a hierarchical Bayesian statistical model. It is estimated by laboratory dating (14C, TL/OSL, AM,...) or by reference dating (i.e. typo-chronology).

- Phases:
  A Phase is a group of Events. It is defined on the basis of archaeological, geological, environmental,... criteria we want to locate in time. Unlike “Event” model, the Phase does not respond to a statistical model: indeed we do not know how Events can be a priori distributed in a phase. However, we may question the beginning, end or duration of a phase from the Events that are observed there (query). A level of a priori information can be added: the Events from one phase may be constrained by a more or less known duration and a hiatus between two phases can be inserted (this imposes a temporal order between the two groups of Events).

- time order relations:
  Events and/or Phases can check order relations. These order relations are defined in different ways: by the stratigraphic relationship (physical relationship observed in the field) or by criteria of stylistic, technical, architectural, etc. development which may be a priori known. These constraints act between facts. A constraint of succession between phases is equivalent to putting order constraints between groups of Events.

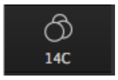
Chapter 2

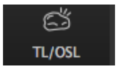
Bayesian modelling

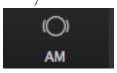
2.1 Event model

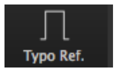
2.1.1 Observations

The measurement may represent :

- a ^{14}C age in radiocarbon 

- a paleodose measurement in luminescence 

- an inclination, a declination or an intensity of the geomagnetic field in archeaomagnetism 

- a typochronological reference (for instance, an interval of ceramic dates) 

- a Gaussian measurement with kown variance 

If needed, these measurements M_i may be converted by ChronoModel into calendar dates using appropriate calibration curves (See section 2.1.5.1).

2.1.2 Definition of an event

Let's say that an event is determined by its unknown calendar date θ. Assuming that this event can reliably be associated with one or several suitable samples, out of which measurements can be made, the event model, implemeted in ChronoModel, combines contemporary dates, $t_1, ..., t_n$, with individual errors, $\sigma_1, ..., \sigma_n$, in order to estimate the unknown calendar date θ.

The following equation shows the stochastic relationship between t_i and θ.

$$t_i = \theta + \sigma_i \, \epsilon_i^{CM} \tag{2.1}$$

where $\epsilon_i^{CM} \sim N(0,1)$ for $i = 1$ to n and ϵ_1^{CM}, ..., ϵ_n^{CM} are independent. θ is the unknown parameter of interest and $\sigma_1, ...\sigma_n$ are the unknown standard deviation parameters.

Such a model means that each parameter t_i can be affected by errors σ_i that can come from different sources (See Lanos & Philippe and see [1]).

ChronoModel is based on a bayesian hierarchical model. Such a model can easily be represented by a directed acyclic graph (DAG) [2]. A DAG is formed by nodes and edges. A node can either represent an observation (data) or a parameter, that can be stochastic or deterministic. An edge is a directed arc that represents dependencies between two nodes. The edge starts at the parent node and heads to the child node. This relationship is often a stochastic one (single arc) but it may also be a determinist one (double arc). The DAG can be read as follow, each node of the DAG is, conditionally on all its parent nodes, independant of all other nodes except of its child nodes.

The following DAG is a representation of the event model. Conditionally on θ and on σ_i, that are the parameters of interest, t_i is independent of all other parameters.

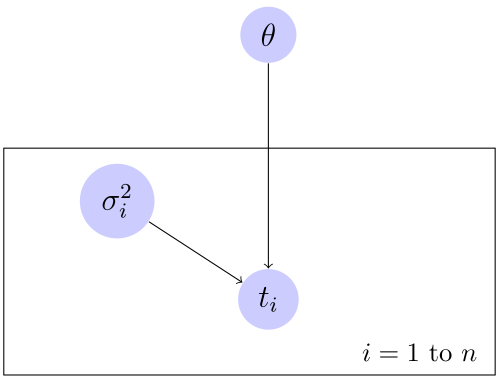


Figure 2.1 – DAG representation of the event model. Directed edges represent stochastic relationships between two variables, blue circles represent model unknown parameters. Rectangular plates are used to show repeated condiotionally indenpendent parameters.

Now we need to define prior information about θ and $\sigma_1, ...\sigma_n$, this is done in the next section. Then the wiggle matching case, specific to radiocarbon datation, is explained. According to the likelihood, three main types of data information may be implemented into ChronoModel: a single measurement with its laboratory error, a combination of multiple measurements or an interval referring to a typo-chronological reference. These differents types are explained in section 2.1.5.

2.1.3 Prior information about θ and $\sigma_1, ...\sigma_n$

Without any other constraint that the beginning, τ_m, and the end, τ_M, of the study period (τ_m and τ_M are be fixed parameters), the unknown calendar date θ is assumed to have a uniform distribution on the study period.

$$p(\theta) = \frac{1}{\tau_M - \tau_m} 1_{[\tau_m, \tau_M]}(\theta) \tag{2.2}$$

The variances σ_i^2, for $i = 1$ to n, are assumed to have a shrinkage uniform distribution (See [3]).

$$p(\sigma_i^2) = \frac{s_0^2}{(s_0^2 + \sigma_i^2)^2} \tag{2.3}$$

where

$$\frac{1}{s_0^2} = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{\hat{s}_i^2} \tag{2.4}$$

with $\hat{s}_i^2$ = the variance of the posterior distribution of t_i obtained by the individual calibration (See [4]).

2.1.4 Particular case of ^{14}C measurements : the wiggle-matching case

This case is specific to radiocarbon datation. Let's say that we have m radiocarbon datations referring to the unknown calendar date θ shifled by a quantity called δ_i. Then, the stochastic relationship between t_i and θ is given by the following equation:

$$t_i = \theta - \delta_i + \sigma_i \, \epsilon_i^{CM} \tag{2.5}$$

where $\epsilon_i^{CM} \sim N(0, 1)$ for i=1 to n and ϵ_1^{CM}, ..., ϵ_n^{CM} are independent.

δ_i may either be a deterministic or a stochastic parameter. Then $t_i + \delta_i$ follows a normal distribution with mean θ and variance σ_i^2.
If δ_i is stochastic, then its prior distribution function is a uniform distribution on $[d_{1i}, d_{2i}]$.

$$p(\delta_i) = \frac{1}{d_{2i} - d_{1i}} 1_{[d_{2i}, d_{1i}]}(\delta_i) \tag{2.6}$$

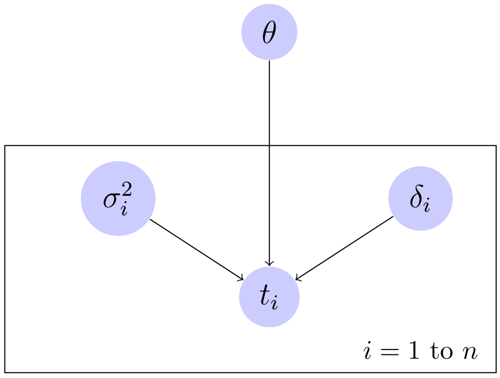


Figure 2.2 – DAG representation of the event model. Directed edges represent stochastic relationships between two variables, blue circles represent model unknown parameters. Rectangular plates are used to show repeated condiotionally indenpendent parameters.

In that case, the associated DAG is presented in Figure 2.2.

2.1.5 Likelihood

As said before, different types of measurement may be included in ChronoModel in order to estimate unknown calendar dates t_i. These different types are the following ones : a ^{14}C age in radiocarbon, a paleodose measurement in luminescence, an inclination, a declination or an intensity of the geomagnetic field in archeaomanetism, a typochronological reference or a Gaussian measurement . Except for the typochronological reference, all other measurement may be associated with a calibration curve. Hence, for the typochronological reference, only the last section may be applied.

2.1.5.1 Calibration curves

- Radiocarbon datations
  Radiocarbon measurements are always reported in terms of years "before present" (BP), that is before 1950. In order to see what a radiocarbon determination means in terms of a true age we need to know how the atmospheric concentration has changed with time. This is why calibration curves are needed.
  Calibration curves implemented in ChronoModel are listed in Table [?]. Other curves or new curves may be used in ChronoModel (See section **??** for more details).

- Archeomagnetism datation
  The process of calibration translates the measured magnetic vector into calendar years. A record of how the Earth's magnetic field has changed over time is required to do this, and is referred to as a calibration curve. A date is obtained

Name	Reference
Uwsy98	??? 1998
IntCal04	Reimer *et al.* 2004 [5]
Marine04	Hughen *et al.* 2004 [6]
ShCal04	McCormac *et al.* 2004 [7]
IntCal09	Reimer *et al.* 2009 [8]
Marine09	Reimer *et al.* 2009 [8]
IntCal13	Reimer *et al.* 2013 [9]
Marine13	Reimer *et al.* 2013 [9]
ShCal13	McCormac *et al.* 2013 [10]

Table 2.1 – Calibration curves implemented in chronoModel

by comparing the mean magnetic vector, defined by the declination and inclination values, with the secular variation curve; the potential age of the sampled feature corresponds to the areas where the magnetic vector overlaps with the calibration curve. Unfortunately, the Earth's magnetic poles have reoccupied the same position on more than one occasion, and can result in multiple age ranges being produced.
A calibrated date is obtained using the separate inclination and declination calibration curves. Probability distributions are produced for the calibrated inclination and declination values, before they are statistically combined to produce a single age estimate. Curves implemented in ChronoModel are listed in Table 2.2.

http://www.brad.ac.uk/archaeomagnetism/archaeomagnetic-dating/

Name	Reference
Gal2002Sph2014_D	? ?
Gal2002Sph2014_I	
Gws2003uni_F	?

Table 2.2 – Calibration curves implemented in chronoModel

- Paleodose datation
  The calibration curve is $g(t) = -t$

- Gaussian datation
  For Gaussian measurements, the calibration curve may is a quadratic polynomial. $g(t) = a * t^2 + b * t + c$. By default the calibration curve is $g(t) = t$.

2.1.5.2 Calibration from one measurement

If the information about t_i come from only one measurement that needs to be calibrated, then the following DAG applies.

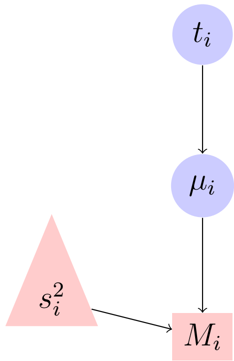


Figure 2.3 – DAG representation of an individual calibration. Directed edges represent stochastic relationships between two variables, blue circles represent model parameters, pink rectangulars nodes represent stochastic observed data, pink triangles represent observed and determinist data.

M_i is the observation data, the measurement made by the laboratory, and s_i^2 is its variance error. In ChronoModel, M_i is assumed to follow a normal distribution with mean μ_i, a latent variable, and with variance s_i^2, the laboratory error. This may be expressed by the following equation :

$$M_i = \mu_i + s_i \, \epsilon_i^{Lab} \tag{2.7}$$

where $\epsilon_i^{Lab} \sim N(0, 1)$.
μ_i may represent for instance the true radiocarbon date or the true archeomagnetism date. We assume that μ_i follows a normal distribution with mean $g_i(t_i)$ and variance $\sigma^2_{g_i(t_i)}$, where g_i is the function of calibration associated with the type of measurement of M_i.

$$\mu_i = g_i(t_i) + \sigma_{g_i(t_i)} \, \epsilon_i^{Cal} \tag{2.8}$$

Hence pooling 2.7 and 2.8 together,

$$M_i = g_i(t_i) + s_i \, \epsilon_i^{Lab} + \sigma_{g_i(t_i)} \, \epsilon_i^{Cal} = g_i(t_i) + S_i \, \epsilon_i^{LabCal} \tag{2.9}$$

where $\epsilon_i^{LabCal} \sim N(0, 1)$ and $S_i^2 = s_i^2 + \sigma^2_{g_i(t_i)}$.
So, conditionally on t_i, M_i follows a normal distribution with mean $g_i(t_i)$ and variance S_i^2.

2.1.5.3 Particular case of ^{14}C measurements : Calibration from multiple measurements

Let's say, we have K measurements M_k from a unique sample. For example, a sample may be sent to K laboratories that give radiocarbon datations. All these measurements refer to the same true radiocarbon date μ. In that case, the bayesian model first gathers all information about μ before calibrating. Hence, $\forall k = 1, ..., K$,

$$M_k = \mu_i + s_k^2\,\epsilon_k^{Lab} \tag{2.10}$$

where $\epsilon_k^{Lab} \sim N(0,1)$ $\forall k = 1, ..., K$ and ϵ_1^{Lab}, ..., ϵ_K^{Lab} are independent. Let's $\overline{M} = \overline{s}^2 \sum_{k=1}^{K} \frac{M_k}{s_k^2}$ and $\overline{s}^2 = \frac{1}{\sum_{k=1}^{K} \frac{1}{s_k^2}}$. Now, as all M_k refer to the same μ, we have

$$\overline{M} = \mu_i + \overline{s}^2 \sum_{k=1}^{K} \epsilon_k^{Lab} \tag{2.11}$$

$$\mu_i = g_i(t_i) + \epsilon^{Cal} \tag{2.12}$$

where $\epsilon^{Cal} \sim N(0, \sigma^2_{g(t)})$. g_i is the function of calibration.
Hence,

$$\overline{M} = g(t) + \overline{s}^2 \sum_{k=1}^{K} \epsilon_k^{Lab} + \epsilon_i^{Cal} = g_i(t_i) + \epsilon^{LabCalMult} \tag{2.13}$$

where $\epsilon^{LabCalMult} \sim N(0, S^2)$ and $S^2 = \overline{s}^2 + \sigma^2_{g_i}(t_i)$. So, conditionally on t_i, the calibrated measurement has a normal distribution with mean $g_i(t_i)$ and variance S_i^2. Figure 2.4 represents the corresponding DAG.

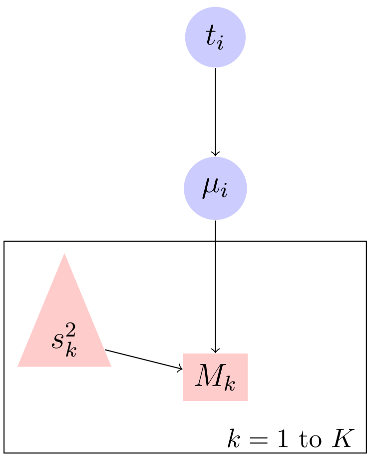


Figure 2.4 – DAG representation of a calibration from multiple measurements. Arrows represent stochastic relationships between two variables, blue circles represent model parameters, pink rectangulars represent stochastic observed data, pink triangles represent determinist observed data.

2.1.5.4 Particularity of archeomagnetism measurements ("combine")

2.1.5.5 Typo-chronological information

Let's say that a typo-chronological information is a period defined by two calendar dates $t_{i,m}$ and $t_{i,M}$, with the constraint $t_{i,m} < t_{i,M}$.
The distribution of $(t_{i,m}, t_{i,M})$ conditional on t_i is given by the following equation

$$p(t_{i,m}, t_{i,M}|t_i) = \lambda^2 e^{-\lambda(t_{i,M}-t_{i,m})} 1_{t_{i,m}<t_i<t_{i,M}} \tag{2.14}$$

where λ is a positive constant. Figure 2.5 represents the corresponding DAG.

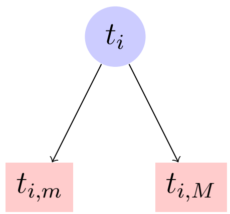


Figure 2.5 – DAG representation of a typo-chronological information. Arrows represent stochastic relationships between two variables, blue circles represent model parameters, pink rectangulars represent stochastic observed data.

2.1.6 Stratigraphic constraints

Several events may be in stratigraphic constraints. Let's say that three events are assumed to happen successively in time, then their true calendar date is assumed to verify the following relationship:

$$\theta_1 < \theta_2 < \theta_3$$

In ChronoModel, constraints links events and not calibrated dates.

Bounds may also be introduced in order to constrain one or several events. Let's say that the three events are assumed to happen after a special event with true calendar date θ^B. Then the following relationship holds.

$$\theta^B < \theta_1 < \theta_2 < \theta_3$$

2.2 Event model including phases

A Phase is a group of Events.

2.2.1 Definition of a phase

A phase is defined on the basis of objective criteria such as archaeological, geological or environmental criteria. We may want to locate a phase in time.
Unlike the event model, a phase does not respond to a statistical model: indeed we do not know how events can a priori be distributed in a phase. Moreover, a phase can only reflect the information given by the events included in that phase. So if a real phase is not entirely covered by archaeological data included is the analysis, the phase as implemented in ChronoModel, will only illustrate the period covered by this data.
However, we may question the beginning, the end or the duration of a phase.

2.2.2 Phase without constraints

If there are neither duration constraints nor hiatus constraints, then the beginning and the end of a phase are simply estimated according to these requests:

2.2.2.1 Estimating the beginning

The beginning of a phase, $\hat{\alpha}$, reflects the minimum of the r events included in the phase:

$$\hat{\alpha} = \min(\theta_{j,\, j=1\ldots r})$$

2.2.2.2 Estimating the end

The end of a phase P is estimated from the highest θ of the r events in the phase:

$$\hat{\beta}^P = \max(\theta^P_{j,j=1\ldots r})$$

2.2.2.3 Estimating the duration

The duration is estimated as the time between the beginning and the end of a phase P:

$$\hat{\tau^P} = \hat{\beta^P} - \hat{\alpha^P}$$

2.2.2.4 Estimating the hiatus between two phases

The hiatus is estimated as the time between the beginning of the next phase $P+1$ and the end of the previous phase P: $\hat{\gamma} = \hat{\alpha}^{P+1} - \hat{\beta}^P$

2.2.3 Information about the duration of a phase

A phase, defined on the basis of for instance archaeological or geological criteria, may be of a fixed duration. Then from the observations, ChronoModel may model events within this duration. Indeed, the duration given is a There are two cases:

- The duration is known : $\tau = \tau_{fixed}$
- The duration is known through a range of values: $\tau \in [\tau_{min}, \tau_{Max}]$.

Consequently, the events θ_j in a phase P have to respect this duration constraint as follow:

$$\max(\theta_j^P) - \min(\theta_j^P) \leq \tau$$

2.2.4 Constraint of hiatus between two phases

The Events from two phases may be constrained by a hiatus inserted between the two phases. This imposes a temporal order between the two groups of Events. There are two cases:

- Hiatus is known : $\gamma = \gamma_{fixed}$
- Hiatus is approximately known in a range : $\gamma \in [\gamma_{min}, \gamma_{Max}]$.

Consequently, the difference between the events θ_j of the two phases P and $P+1$ have to respect this constraint:

$$\max(\theta_j^{P+1}) - \min(\theta_j^P) \geq \gamma$$

Chapter 3

Numerical methods

In general, the posterior distribution does not have an analytical form. Elaborated algotithms are then required to approximate this posterior distribution.
Markov chain Monte Carlo (MCMC) is a general method based on drawing values of θ from approximate distributions and then corrected those draws to better approximate the target posterior distribution $p(\theta|y)$. The sampling is done sequentially, with the distribution of the sampled draws depending on the past value drawn. Indeed, a Markov chain is a sequence of random variables $\theta^(1)$, $\theta^(2)$, ..., for which, for any t, the distribution of $\theta^(t)$ given all previous θ's depend only on the recent value, $\theta^{(t-1)}$ [11, 12].

3.1 Choice of the MCMC algorithm

A convenient algorithm useful in many multidimensional problems is the Gibbs sampler (or conditional sampling) [11, 12].
Let's say we want to approximate $p(\theta_1, \theta_2, ..., \theta_d|y)$. The algorithm starts with a sample of initial values $(\theta_1^{(0)}, \theta_2^{(0)}, ..., \theta_d^{(0)})$ randomly selected . The first step of the algorithm is to update the first value by sampling a candidate value of $\theta_1^{(1)}$ knowing $\theta_2^{(0)}, ..., \theta_d^{(0)}$ from the full conditional distribution $p(\theta_1^{(0)}|\theta_2^{(0)}, ..., \theta_d^{(0)})$. The next step is to find a candidate value $\theta_2^{(1)}$ knowing $\theta_1^{(1)}, \theta_3^{(0)}, ..., \theta_d^{(0)}$ using the full conditional distribution $p(\theta_2^{(0)}|\theta_1^{(1)}, \theta_3^{(0)}, ..., \theta_d^{(0)})$.. And so on... Then the dst step is to find a candidate value for $\theta_d^{(1)}$ knowing $\theta_1^{(1)}, \theta_2^{(1)}, ..., \theta_{d-1}^{(1)}$. This process is then iteratively repeated.

Starting values : In ChronoModel, the initial values of each Markov chain are randomly selected. More details ar egiven in the appendix.

In ChronoModel, two main algorithms are implemented, the **rejection sampling**

method and the **Metropolis Hastings algorithm**. Both algorithms required a proposal density function, that should be easily sampled from, in order to generate new candidate values. For the rejection sampling algorithm it is common to use, if possible, the prior function or the likelihood as a proposal function. For the Metropolis-Hastings algorithm, a common choice is to use a symmetric density, such as the Gaussian density.
Depending on the type of the parameter, the event, the mean of a calibrated measure, the variance of a calibrated measure or a bound, different methods are proposed in order to generate new candidate values at each step of the Gibbs sampler. These methods are described here in turn.

3.1.1 Drawings from the conditional posterior distribution of the event θ

Three different methods can be chosen.

- **Rejection sampling with a Gaussian proposal** [13]
- **Rejection sampling with a Double exponential proposal** [13]
- **Metropolis-Hastings algorithm with an adaptative Gaussian random walk** [14]

The first two methods are exact methods. We recommend to use one of these two methods except when the event is involved in stratigraphic constraints. In that case the last method should rather be used.

3.1.2 Drawings from the full conditional posterior distribution of the calibrated date t_i

In this case, three different methods can be chosen.

- **Metropolis-Hastings algorithm using the posterior distribution of calibrated dates** ($P(M_i|t_i)$)
  This method is adapted for calibrated measures, namely radiocarbon measurements or archeomagnetism measurements but not for typo chronological references, and when densities are multimodal.

- **Metropolis-Hastings algorithm using the parameter prior distribution** ($P(t_i|\sigma_i^2, \theta)$)

This method is recommended when no calibration is needed, namely for TL/OSL, gaussian measurements or typo-chronological references.

- **Metropolis-Hastings algorithm using an adaptative Gaussian random walk**
  This method is recommended when no calibration is needed or when there are stratigraphic constraints. This method is adapted when the density to be approximated is unimodal.

3.1.3 Drawings from the conditional posterior distribution of the variance of a calibrated measure σ_i^2

In this case, only one method is implemented in ChronoModel, the uniform skrinkage as explained in Daniels [3]. The full conditional density is unimodal, hence the Metropolis Hastings algorithm can be implemented here. The proposal density involved is an adaptative Gaussian random walk [14]. The variance of this proposal density is adapted during the process.

3.2 MCMC settings

The algorithms described above generate a Markov chain for each parameter, that is a sample of values out of the posterior full conditional distribution. Now, wait for all the Markov chains to reach equilibrium. Let's say this occurs at some time T. The time before T is usually called the burn-in period.

3.2.1 Burn

This period is used to "forget" the initial values randomly selected for each parameter.

3.2.2 Adapt

This is a period needed to calibrate all variances of adaptative Gaussian random walks. This method is at least used to draw values from the individual variances σ_i^2. The variance is calculated on all iterations of the first batch and the acceptation rate is estimated. The adaptation period goes on with another batch using the variance estimated on this last batch unless the acceptation rate calculated is included between 40% and 46% or the maximum number of batches is reached.

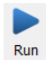
3.2.3 Acquire

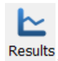
In this period, all Markov chains are assumed to have reached their equilibrium distribution. Of course, this has to be checked and the next section provides useful tools that can help controlling if the equilibrium is actually reached. If so, Markov chains may be sampled and information about conditional posterior distribution can be extracted.

Sampling from these Markov chains need to be carefully made. Indeed, successive value of a Markov chain are not independent. In order to limit the correlation of the sample, we can choose to thin the sample by only keeping equally spaced values.

Chapter 4

Results and Interpretations

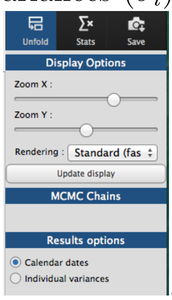
After having gone through the running process 

, the results tab appears 

. Now, before any interpretations, the Markov chains have to be checked.

4.1 Checking the Markov chains

When Markov chains are generated, two points have to be verified : the convergence of the chains and the absence of correlation between successive values. If the Markov chain has not reach its equilibrium, values exctracted from the chains will give inapropriate estimates of the posterior distribution. If high correlation remains between successive values of the chain, then variance of the posterior distribution will be biased. Here are some tools to detect whether a chain has reach its equilibrium and whether successive values are correlated. We also give indications about what can be done in these unformtunate situations.

By default, results correspond to the dates density functions. To see results regarding individual variances (σ_i), click on the "Individual variances" under the "Results option" section 

.

4.1.1 Is the equilibrium reached ? Look at the history plots (History plot tab)

Unfortunatly, there is no theoretical way to determine how long will be the burn-in period of a Markov chain. The first thing to do is to observe the trace (History plots) of the chain and inspect it for signs of convergence. Traces should have good mixing properties, they should not show tendancies or constant stages. Figure 4.1 displays an example of good mixing properties. History plots of dates and variances should be checked.

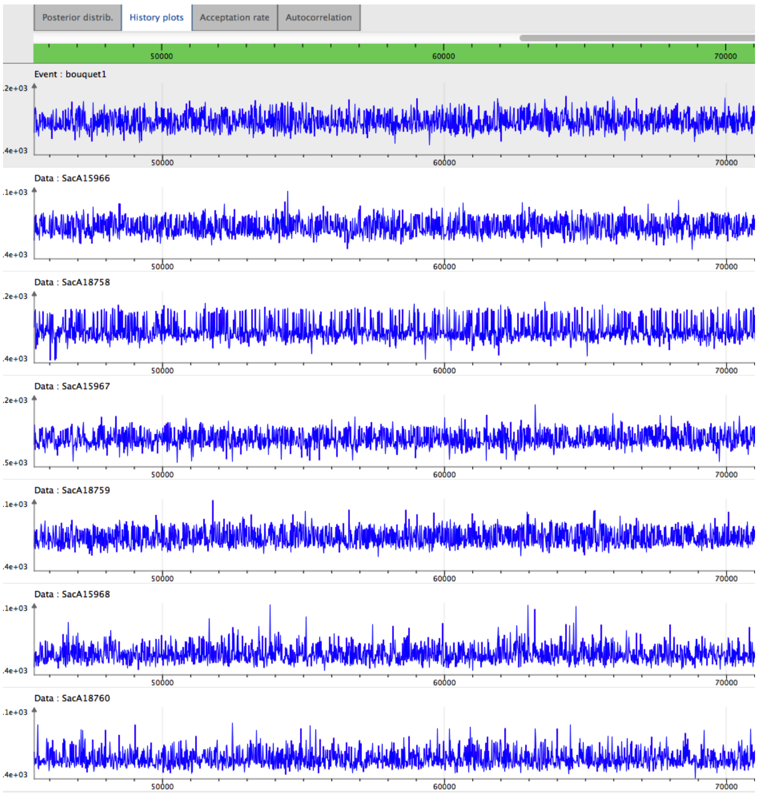


Figure 4.1 – Examples of history plots with good mixing propoerties

Producing parallel Markov chains, all with different starting values, can help deciding if (and when) a chain has reach its equilibrium. In that case, the posterior ditribution of each chain might be overlayed. If the equilibrium is reached, then the posterior distrubtions should be similar.

What should be done when equilibrium is not reached ? First, MCMC settings should be changed by asking for a longer burn-in period or a longer adapt period. Then, if the equilibrium is still not reached, changing the algorithm used to draw from full conditional posterior distributions might be of help.

4.1.2 Correlation between successive values ? Look at the autocorrelation functions (Autocorrelation tab)

A Markov chain is a sequence of random variables $\theta^{(1)}$, $\theta^{(2)}$, ..., for which, for any t, the distribution of $\theta^{(t)}$ given all previous θ's depend only on the recent value, $\theta^{(t-1)}$ [11, 12]. Hence, a high correlation between two consecutive values is expected, but not between all values. To check whether the chain is highly correlated, observe the autocorrelation plot. Only the first correlations should be high, the remaining correlations should be negligible. Autocorrelations should have an exponential decrease. Autocorrelations of dates and of variances should be checked.

What should be done if correlation is high ? A good thing to do is to increase the Thinning interval and the sample the chain by keeping values with higher intervals between them.

4.1.3 Look at the acceptation rates (Acceptation rates tab)

The Metropolos Hastings algorithm generates a candidate value from a proposal density. This candidate value is accepted with a probability. An interesting point is the acceptation rate of this proposal density. The theoretical optimal rate is 43% [14].
In ChronoModel, this algorithm is at least used for each individual variance and for calibrated dates but may also be used for events.
Figure 4.3 displays an example of acceptation rates that are close to 43%. Acceptation rates of dates and of variances should be checked.

What should be done if the acceptation rates are not close to 43% ?

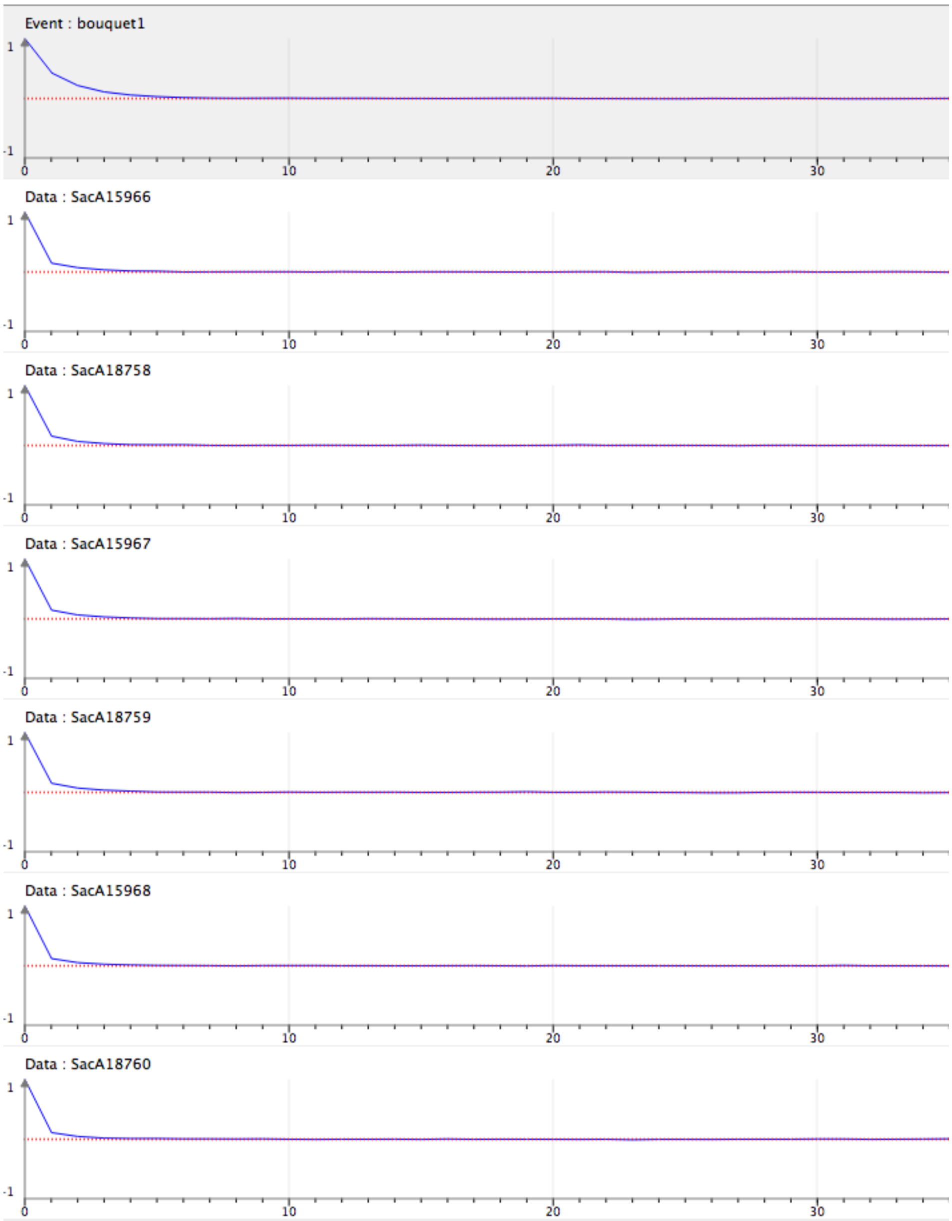


Figure 4.2 – Examples of autocorrelation functions that fall quickly under the 95% confidence interval

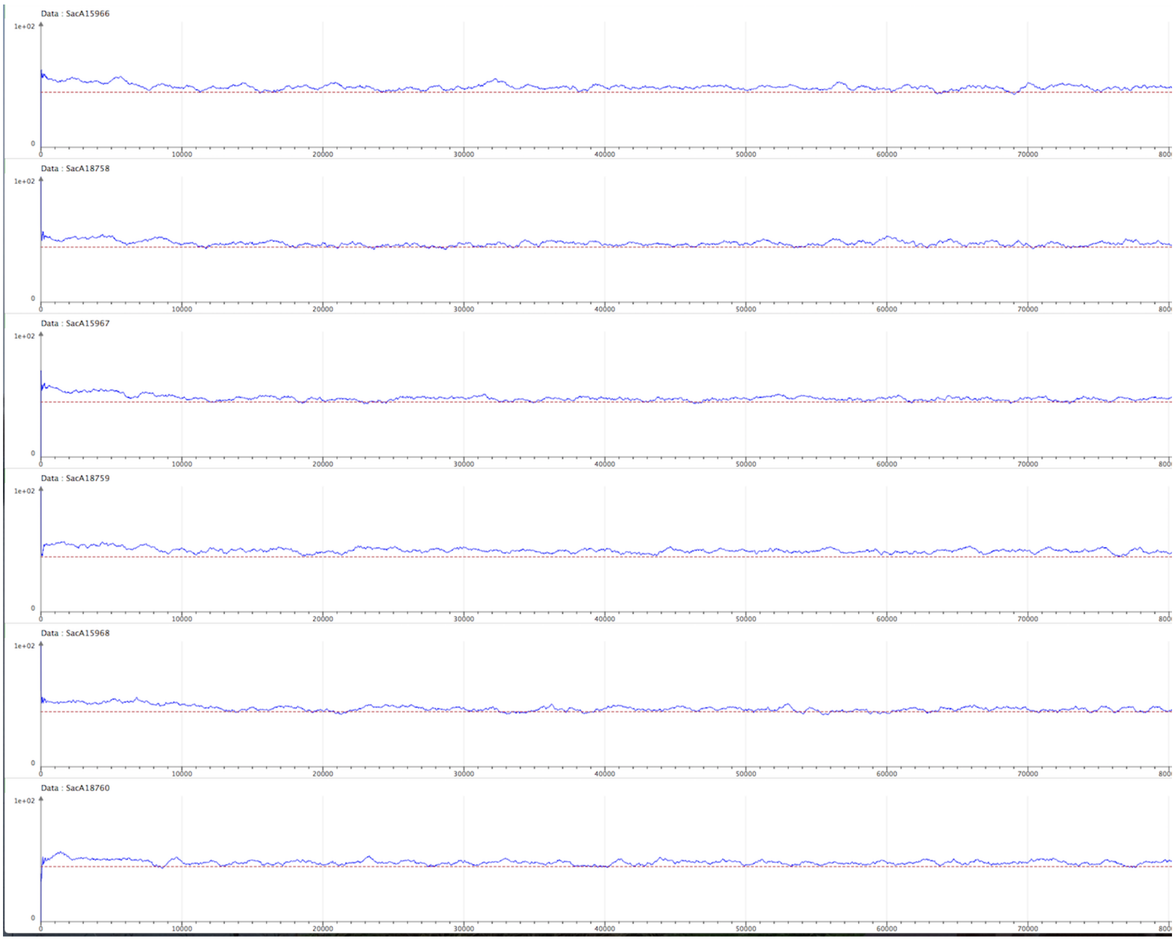


Figure 4.3 – Examples of acceptation rates close to 43%

4.2 Interpretation

If all Markov chains have reached their equilibrium and are not autocorrelated, then the statistical results have a real meaning.

4.2.1 Calibrated curves

4.2.2 Posterior density

Posterior densities are in fact marginal densities. These densities should be interpreted parameter by parameter.

4.2.2.1 Event and calendar dates

If the densities of calibrated dates associated with the same event seem to gather about the same period of time, then the modellisation is correct.

4.2.2.2 Individual variance

Variances density functions should be small. If not, the information given by the measurement associated with this wide variance density function has a low impact on the posterior distribution of the event. It is the case of outliers (See the example of Sennefer's tomb, Bouquet 2).

4.2.3 Statistical results

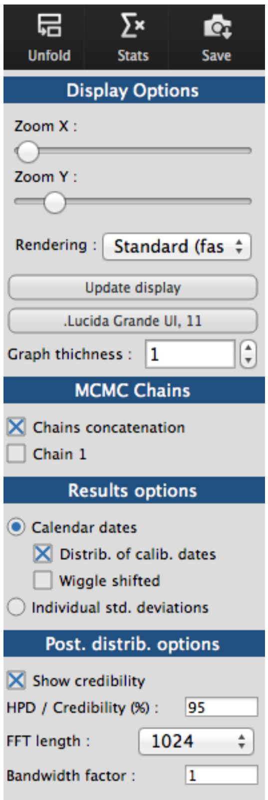
ChronoModel gives also a list of statistical results. Click on  to see those results.

- Maximum a posteriori (MAP) : the highest mode of the posterior density
- Mean : the mean of the posterior density function
- Std deviation : the standard deviation of the data
- Q1 : the numerical value separating the lower 25% of the data from the higher 75%
- Q2 : the numerical value separating the lower 50% of the data from the higher 50%
- Q3 : the numerical value separating the lower 75% of the data from the higher 25%

- Crediblity interval (CI) or Bayesian confidence interval: the smallest credible interval
- Highest probability density region (HPD) : the region with the highest probability density

CI and HPD regions may be of the chosen percentage. By default, the 95% intervals are estimated. To change that, go on the Posterior Distribution tab 

and



change the percent from the options on the right handside of the window.

Chapter 5

Examples

5.1 Scenario

In this example, we present a fictitious archaeological excavation with stratigraphy on several structures. We have only one 14C specimen on each.

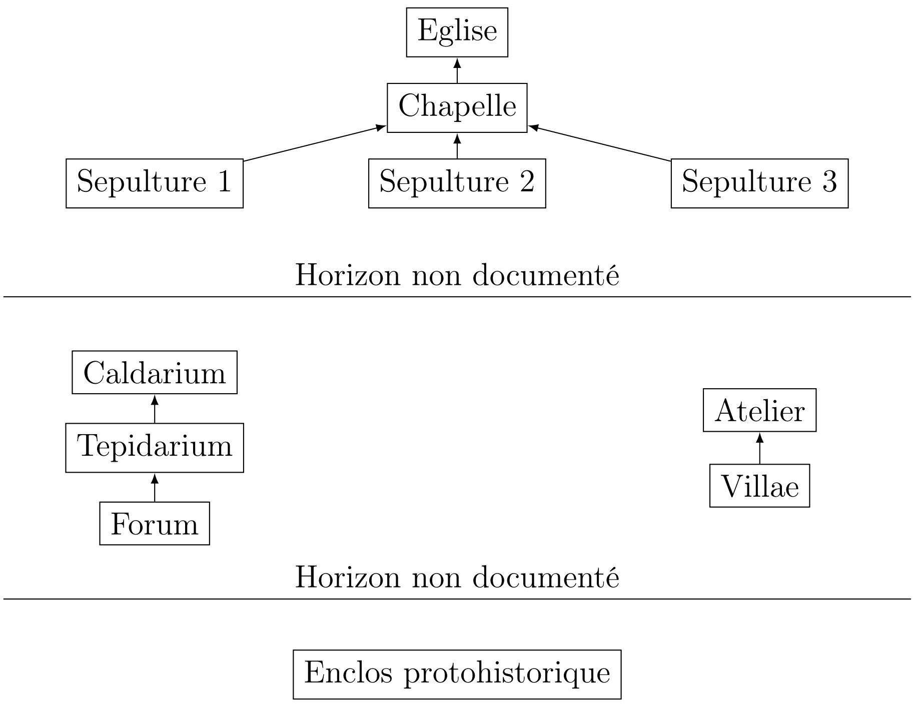


Figure 5.1 – Field model

The following table gives the corresponding measurements vs structures.

Structure	**Ref. 14C**	**Value and error**
Enclos	Pr1	2540 ± 50
Villae	GR5	1850 ± 30
Atelier	GR4	1735 ± 30
Forum	GR3	1764 ± 30
Tepidarium	GR2	1760 ± 30
Caldarium	GR1	1734 ± 30
Sepulture 1	M3	1350 ± 35
Sepulture 2	M4	1390 ± 30
Sepulture 3	M5	1370 ± 50
Chapelle	M2	1180 ± 30
Eglise	M1	950 ± 35

5.1.1 Only one sequence

First chronological model is putting each measurement in an Event for each structure and drawing constraints between these Events, as showing in the figure 5.2

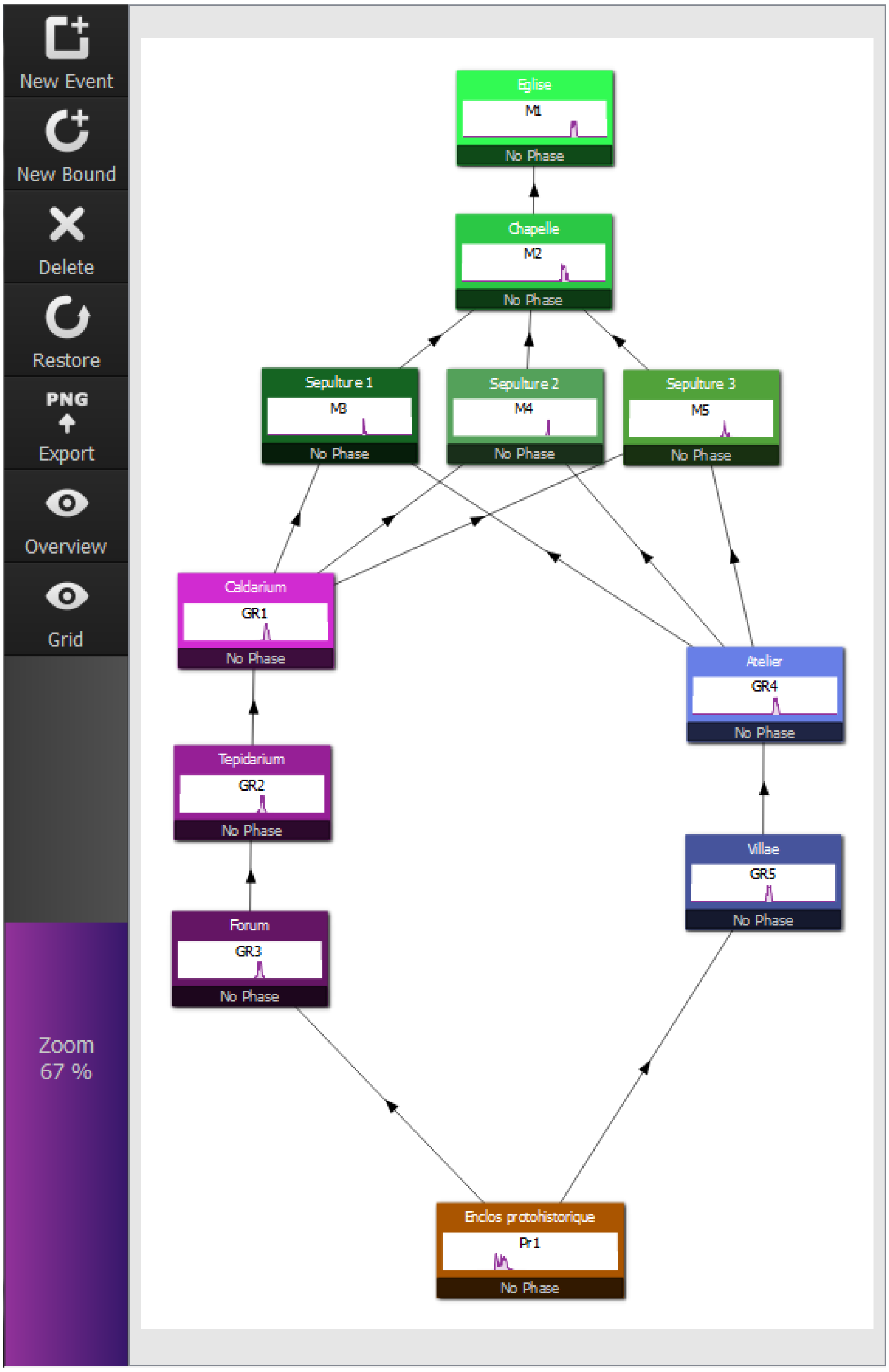


Figure 5.2 – Sequential model

After computing by clicking on the run button, the screen toggle to the result window.

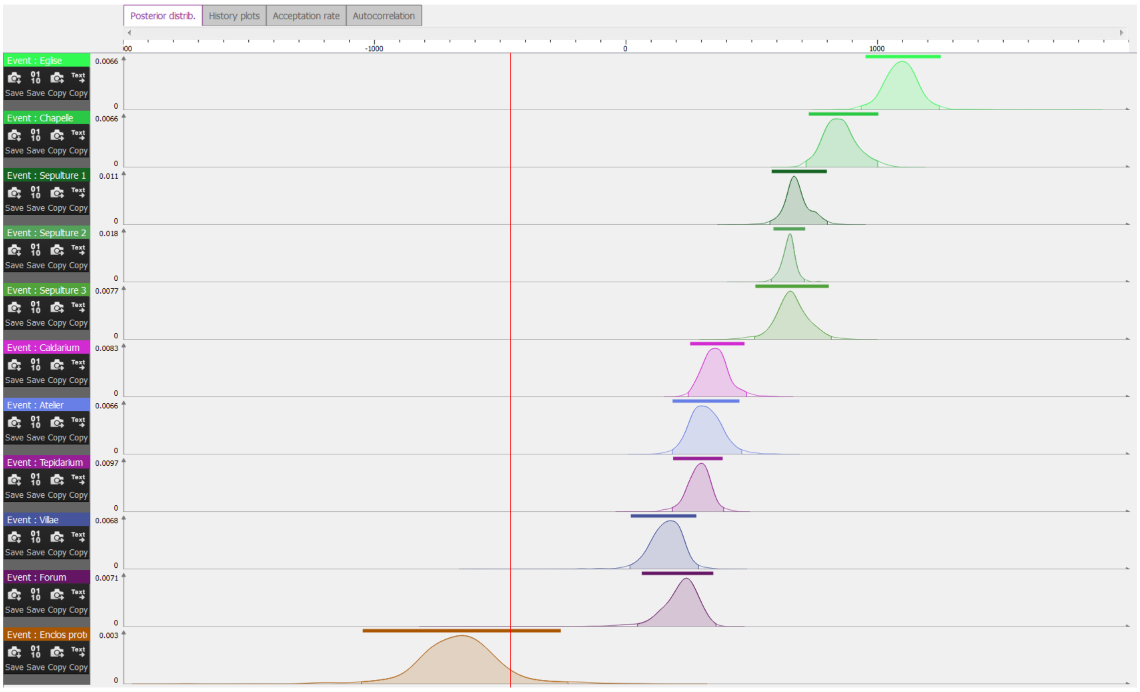


Figure 5.3 – Result window for the sequential model

5.1.2 Several sequences and phases

The other equivalent way to build the chronology is to introduce phases. In our example, we can see 4 sequences nested in 4 phases. Each phase corresponds to a group structures (model figure 5.4)

After computing, the screen toggles to the result window. So, we can see the same result as the sequential model. But now, we can get a result window with all phases.

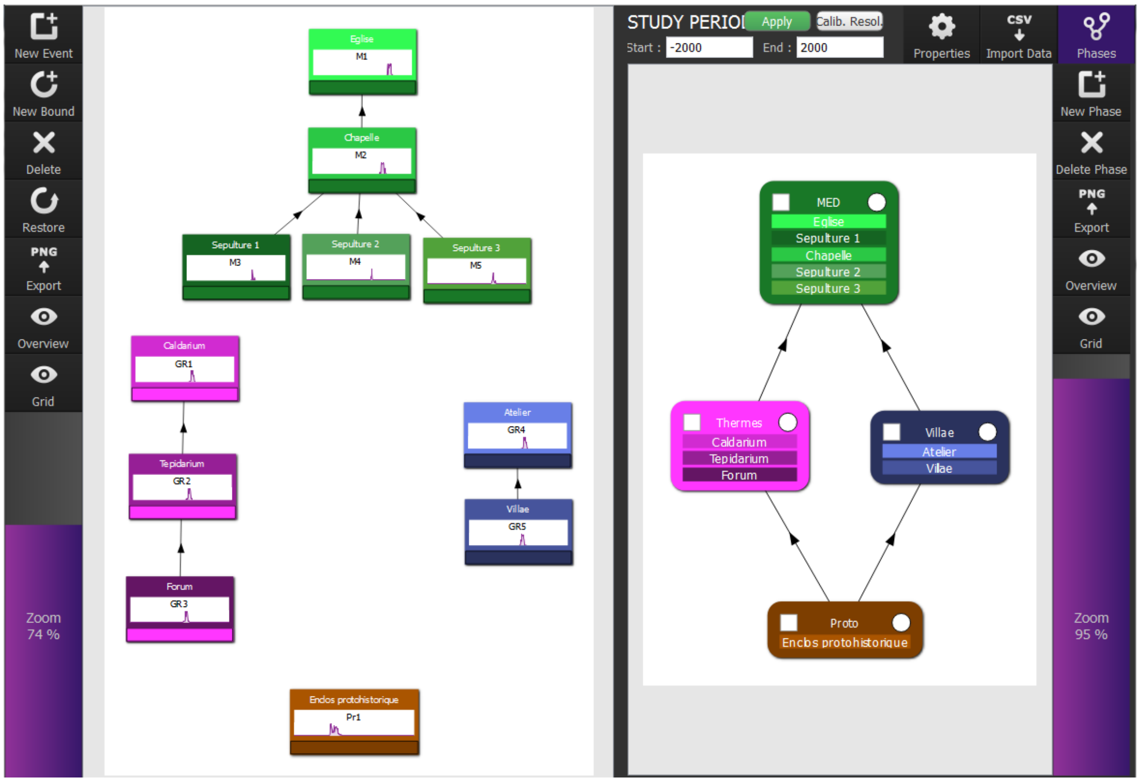


Figure 5.4 – Event and phase model

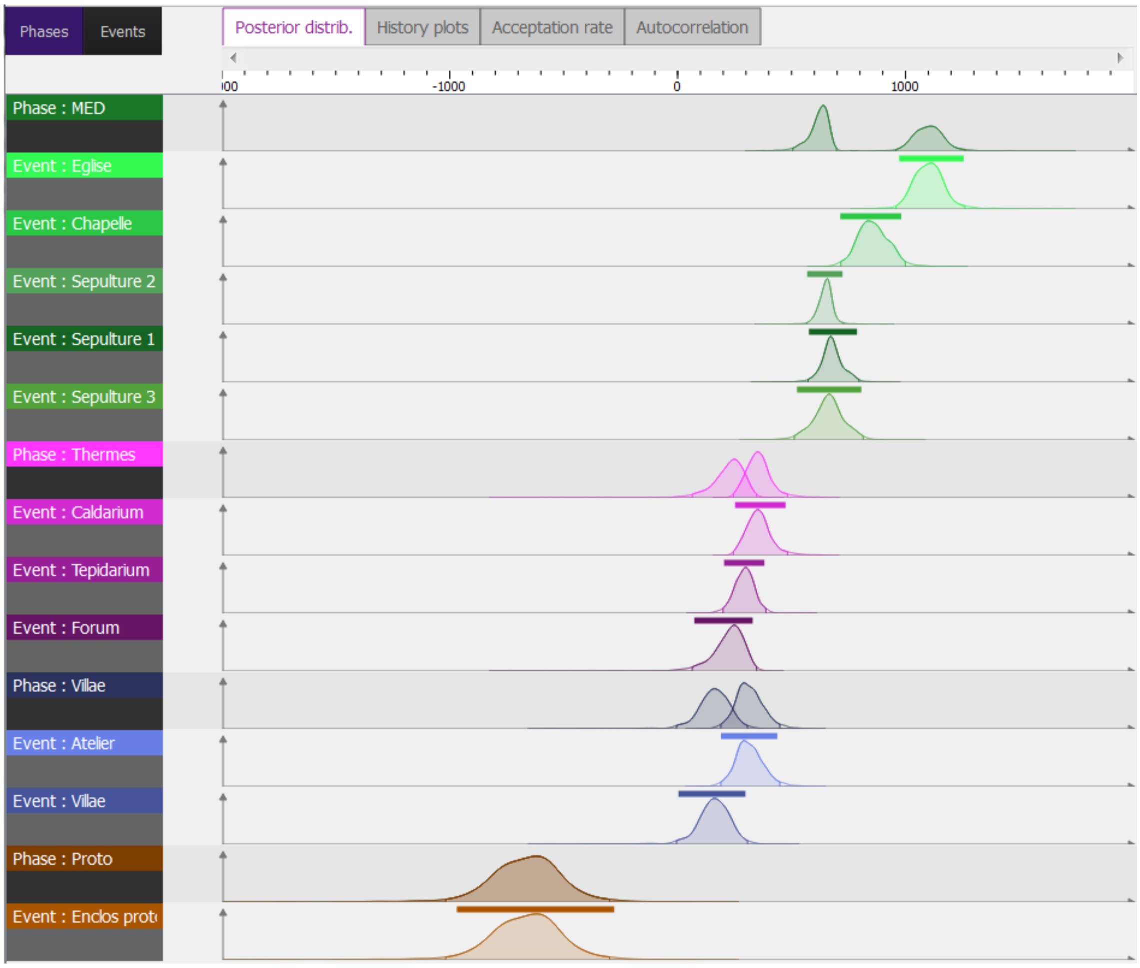


Figure 5.5 – Events and phases results

5.1.3 Several sequences and phases

Now, keeping the same sequences, we add new phases corresponding to given typo-chronological criteria.

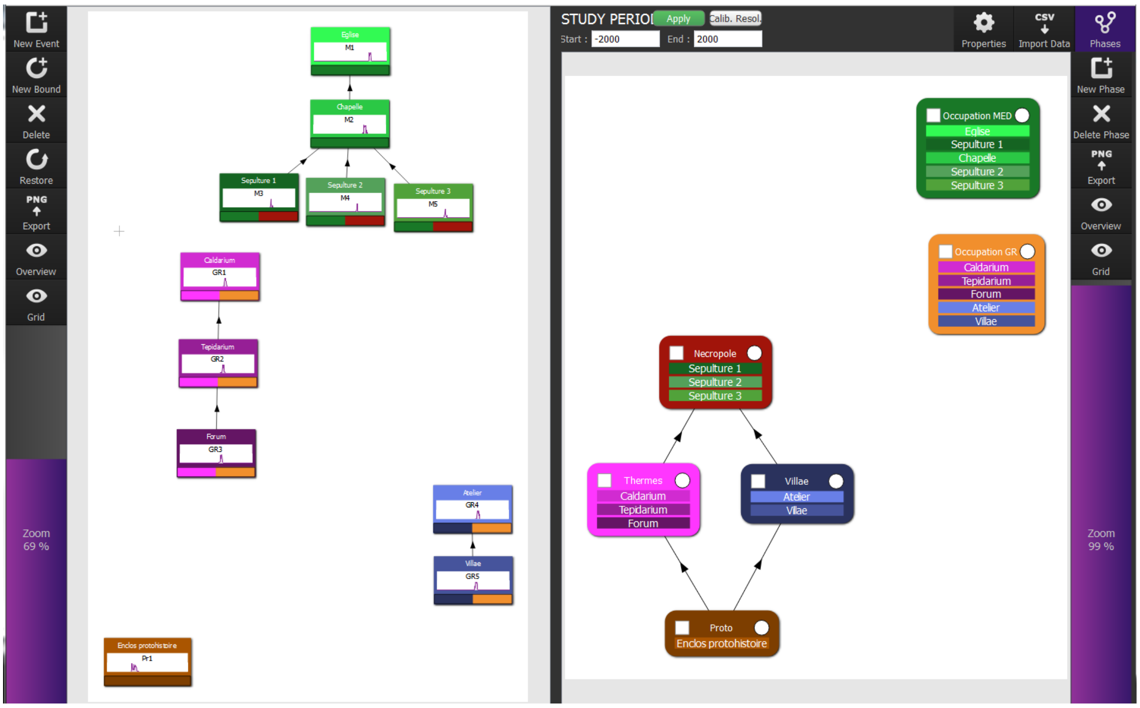


Figure 5.6 – Events, squences and phases model

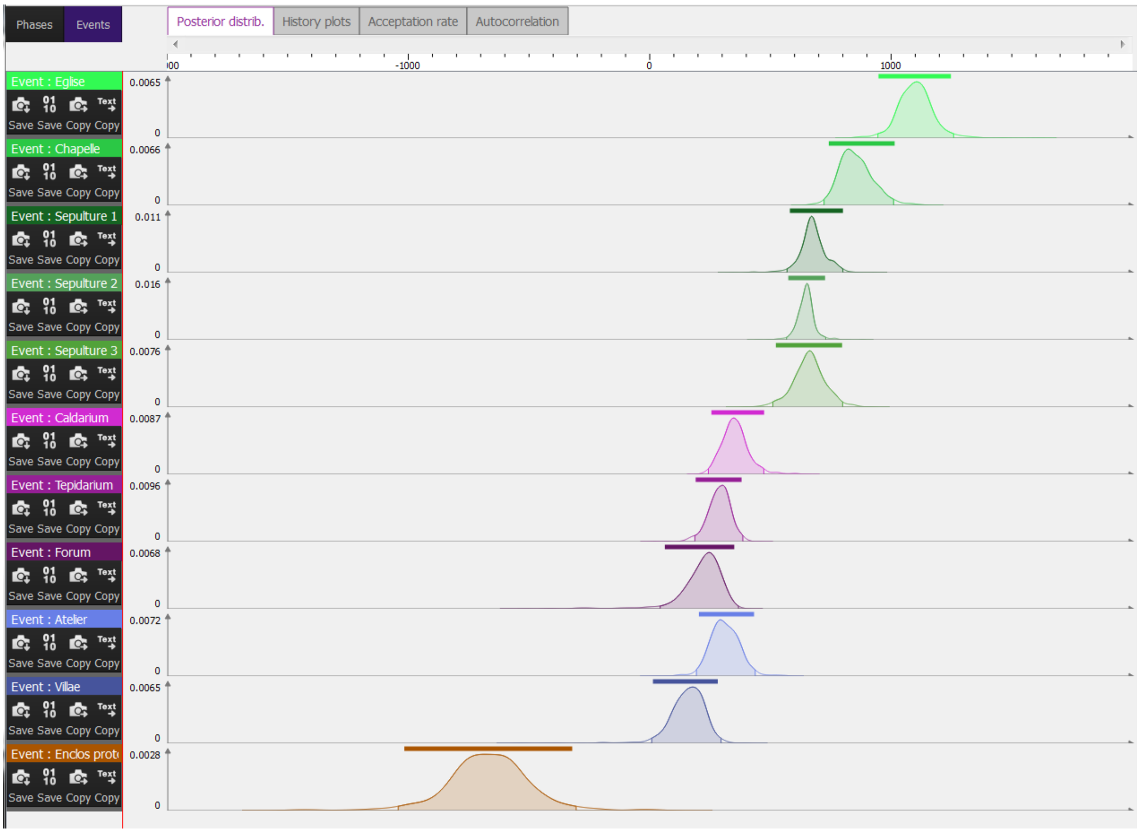


Figure 5.7 – Events results

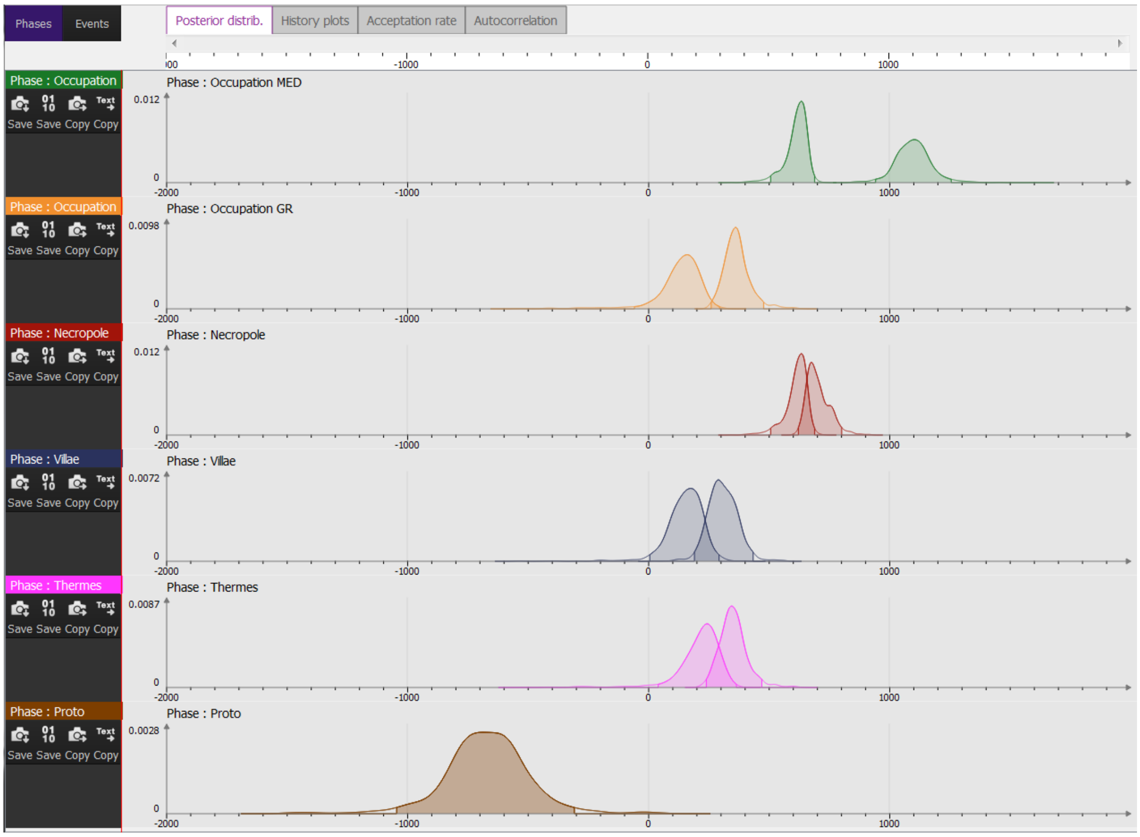


Figure 5.8 – Phases results

5.2 Radiocarbon datation in Sennefer's tomb (Egypt)

We used data published in the article of Anita Quiles [15]. Several bouquets of flowers were found in Sennefer's tomb at Deir el-Medineh. As they were found at the entrance of the tomb, they were assumed to date the same archeological event: the burial of Sennefer. The objective is to date this burial event using ChronoModel.

Samples were extracted from different short-lived plants (leaves, twigs, etc) on each bouquet in order to ensure the consistency of the dates. All samples were radiocarbon dated.

5.2.1 Bouquet 1

Let's say we want to estimate the calendar date of bouquet 1. 6 samples were extracted from Bouquet 1 and radiocarbon dated (Bouquet1.CSV may be dowloaded from the website). The modelisation of this bouquet by ChronoModel is represented by Figure 5.9.

Each radiocarbon measurement is calibrated using IntCal09 curve. No reservoir offset is taken into account. The study period is chosen to start at -3000 and end at 0 using a step of 1 year.

The method used to draw values from the conditional posterior distribution of the event Bouquet 1 is the default one, the rejection sampling using a double exponential proposal. The method used to draw values from the conditional posterior distribution of the datations is also the default one: Metropolis-Hastings algorithm using the posterior distribution of calibrated dates.

We start with 1 000 iterations in the Burn-in period, 1 000 iterations in each of the 100 maximum batches in the Adapt period and 100 000 iterations in the Acquire period using thinning intervals of 1. Only one chain is produced.

Figure 5.10 represents the marginal posterior densities of each date parameter (the event and the calendar dates of the calibrated measurements). In this example, 95% intervals (CI and HPD) are represented. We can see that all calendar dates seem to be contemporary dates. Numerical values, displayed in Figure 5.12, show that the MAP and the mean values were quite close, as well as HPD et CI intervals. The event is dated with at -1370 (mean value) associated with its 95% HPD interval [-1417; -1314]. Figure 5.11 shows the history plots (or the trace of the Markov chains) of each date parameter. During the acquisition period, all chains seem to have good mixing properties. We may assum that all chains have reach their equilibrium be-

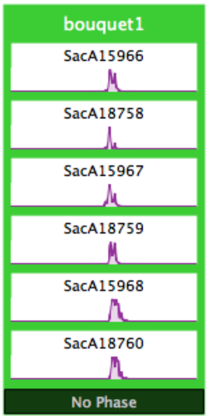


Figure 5.9 – Modelisation of Bouquet 1 with ChronoModel

fore the acquisition period. Figure 5.13 presents the acceptation rates of each date parameter. All rates are close to the optimal rate of 43%. Figure 5.14 displays the autocorrelation function of each date parameter. We can see that all autocorrelation functions decrease exponentially and fall under the 95% confidence interval after a lag of 30 for calibrated dates and after a lag of 50 for the event. This autocorrelation between successive values may be reduced by increasing the thinning interval at 10 for example. In order to keep 100 000 observations in the acquire period, we ask for 1 000 000 but only 1 out of 10 values were kept for the analysis. The autocorrelation functions obtained decrease exponentially and fall under the 95% confidence interval after a lag of 6 for each parameter. However, with this new MCMC settings, all other results are similar to those already given.

Now, let's look at the individual standard deviations results. The marginal posterior densities of each individual standard deviations, presented Figure 5.16, seem to be of similar behavior, with a mean about 50 and a standard deviation about 48 (numerical values displayed in Figure 5.17). History plots of these individual standard deviations, presented Figure 5.18, seem to have good mixing properties. Hence, the equilibrium is assumed to be reached. Each acceptation rates, presented Figure 5.19, are close to the optimal rate of 43%. And finally, each autocorrelation function, displayed in Figure 5.19, shows an exponential decrease and all values fall under the 95% interval of signification after a lag of 10.

In conclusion, the modelisation of Bouquet 1 seems consistent. All individual standard deviations take values close to 50 compared to -1400 for the event. That is to say, standard deviations are rather small compared to the event's posterior mean.

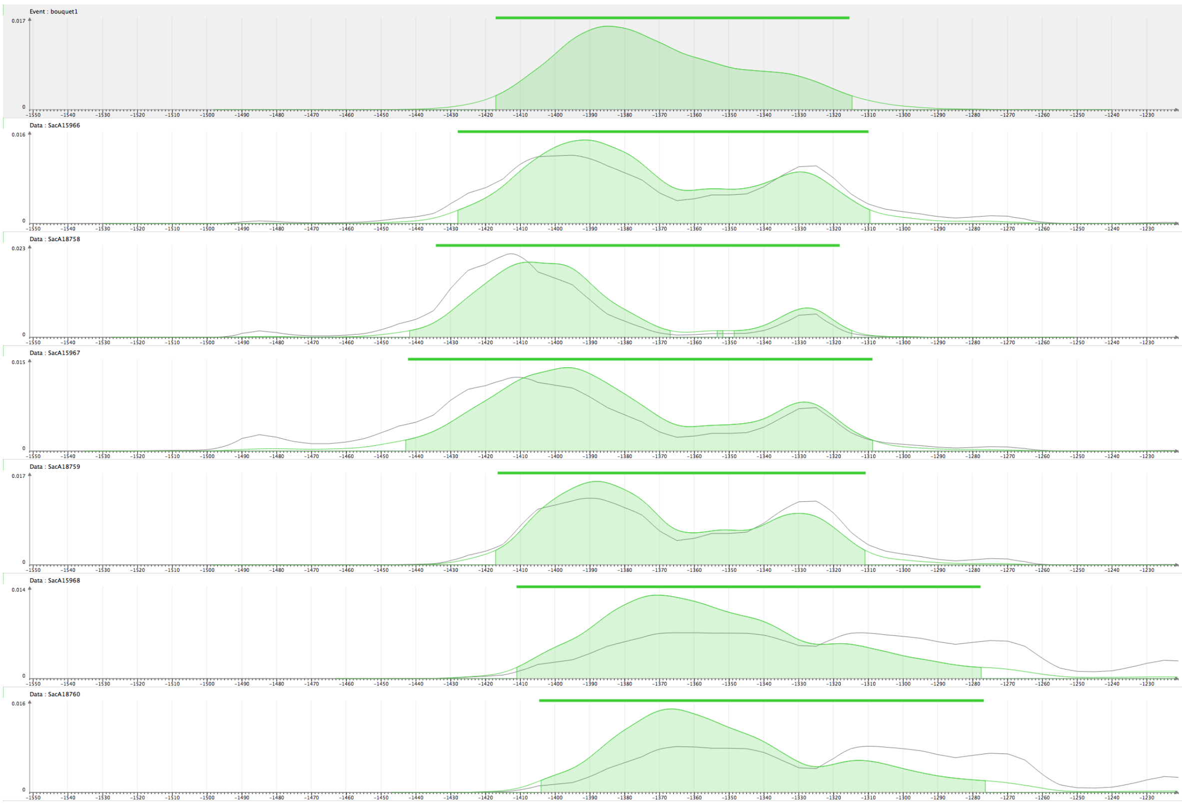


Figure 5.10 – Marginal posterior densities related to the modellisation of Bouquet 1 . The dark lines correspond to distribution of calibrated dates, the green lines correspond to posterior density functions. Highest posterior density (HPD) intervals are represented by the green shadow area under the green lines. Credibility intervals are represented by thick lines drawn above the green lines.

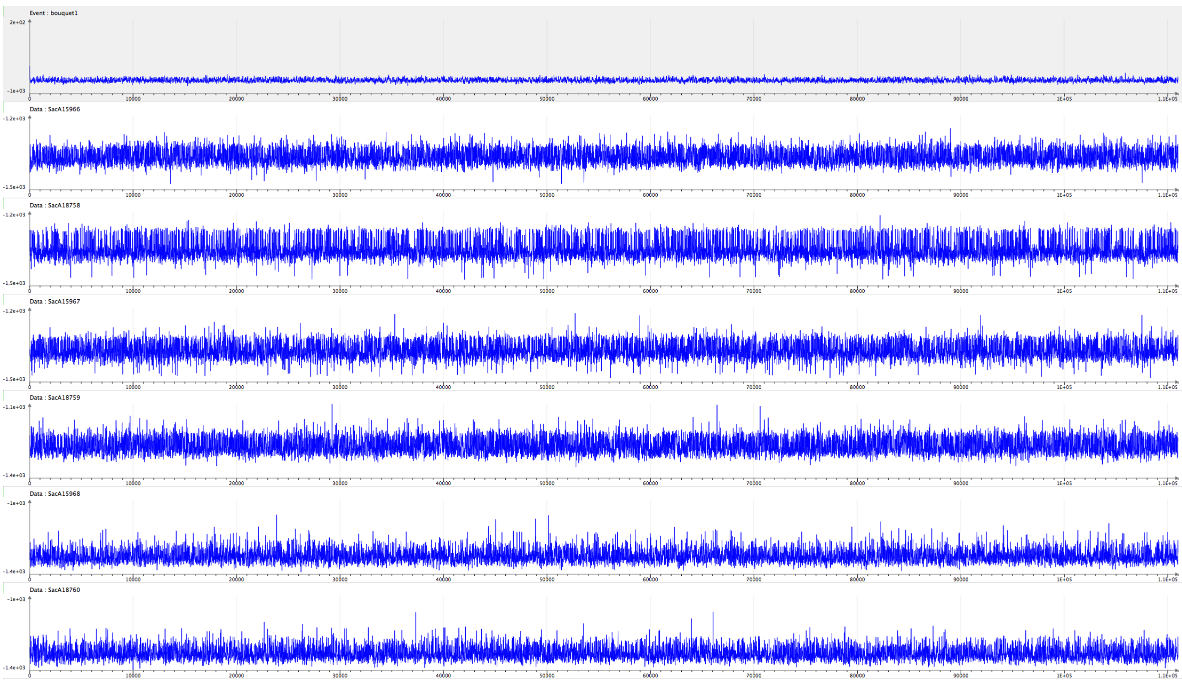


Figure 5.11 – History plots related to the modellisation of Bouquet 1

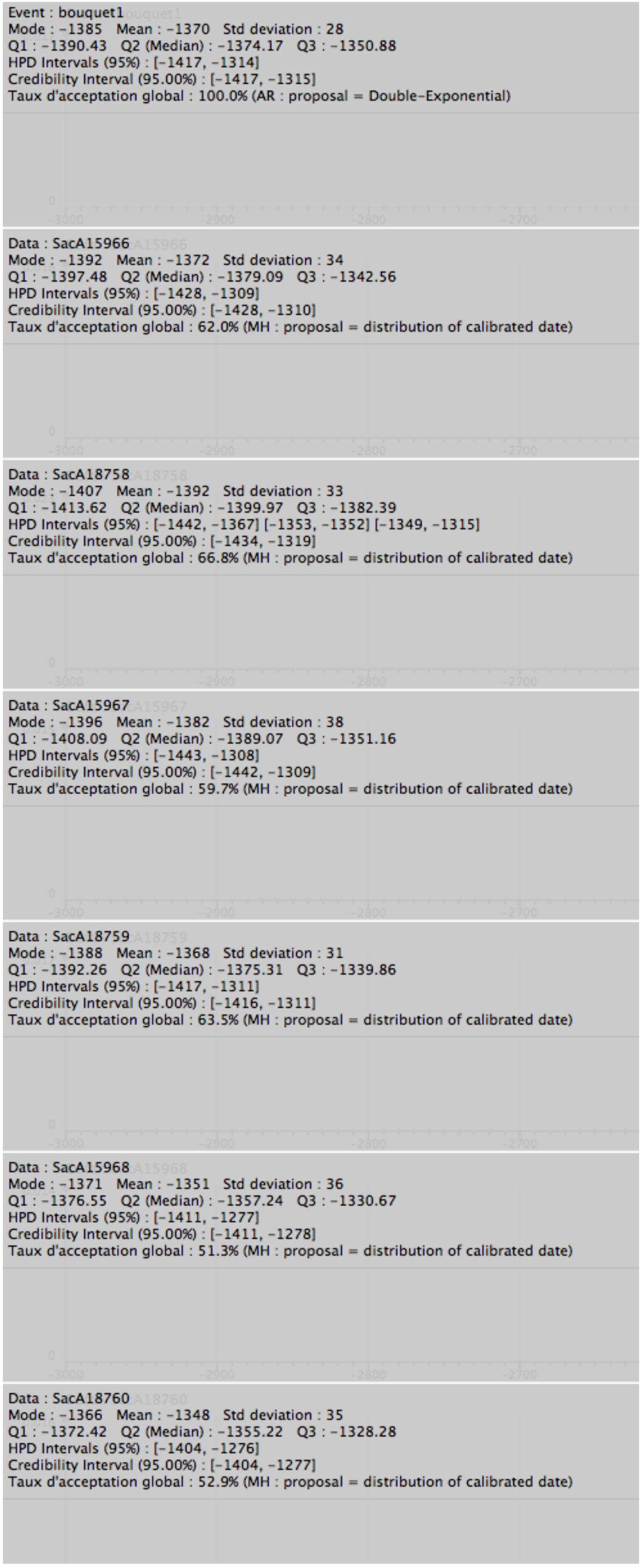


Figure 5.12 – Numerical values related to the modellisation of Bouquet 1

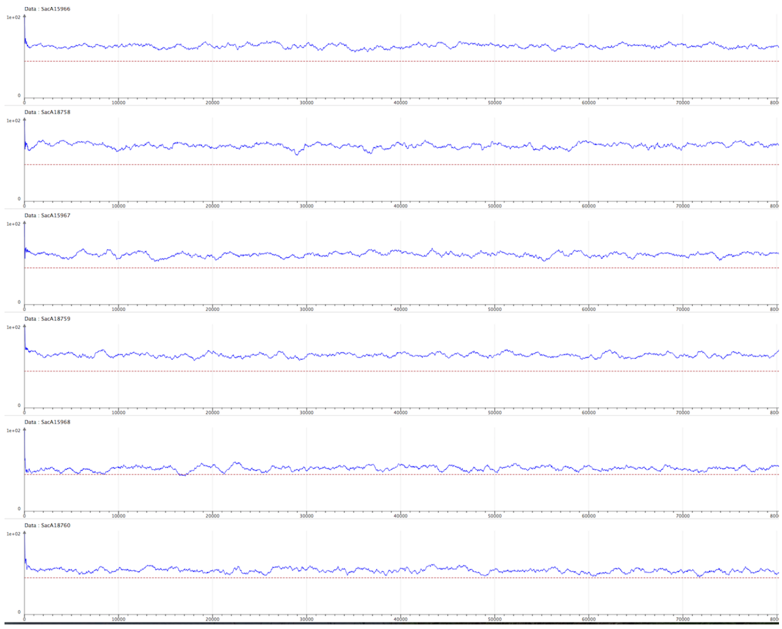


Figure 5.13 – Acceptation rates related to the modellisation of Bouquet 1

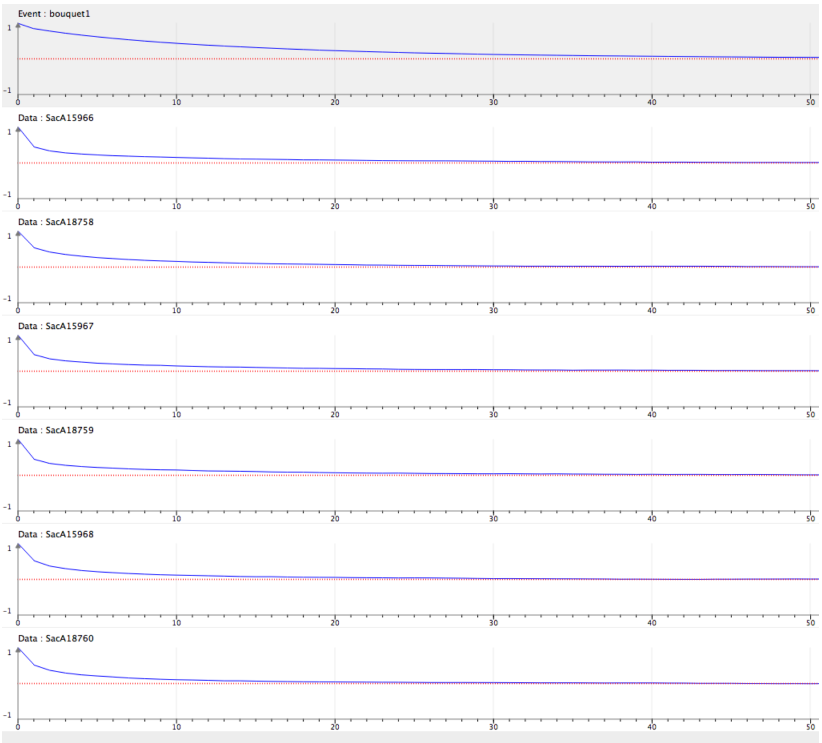


Figure 5.14 – Autocorrelation functions related to the modellisation of Bouquet 1

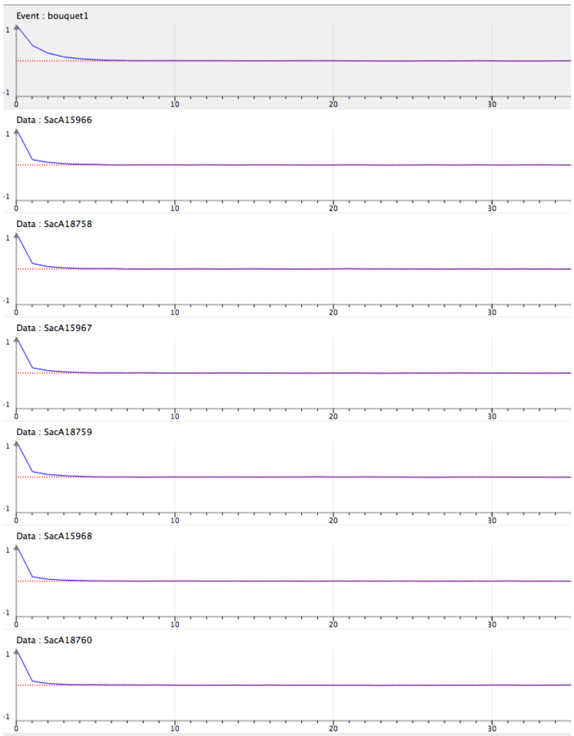


Figure 5.15 – Autocorrelation functions related to the modellisation of Bouquet 1

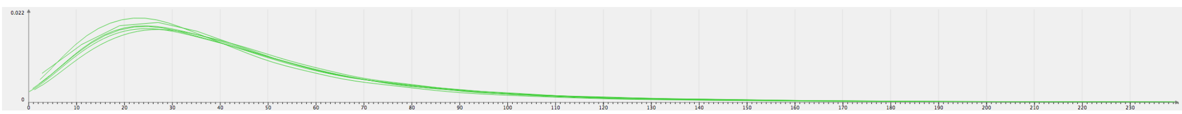


Figure 5.16 – Marginal posterior densities of indivudal standard deviations related to the modellisation of Bouquet 1

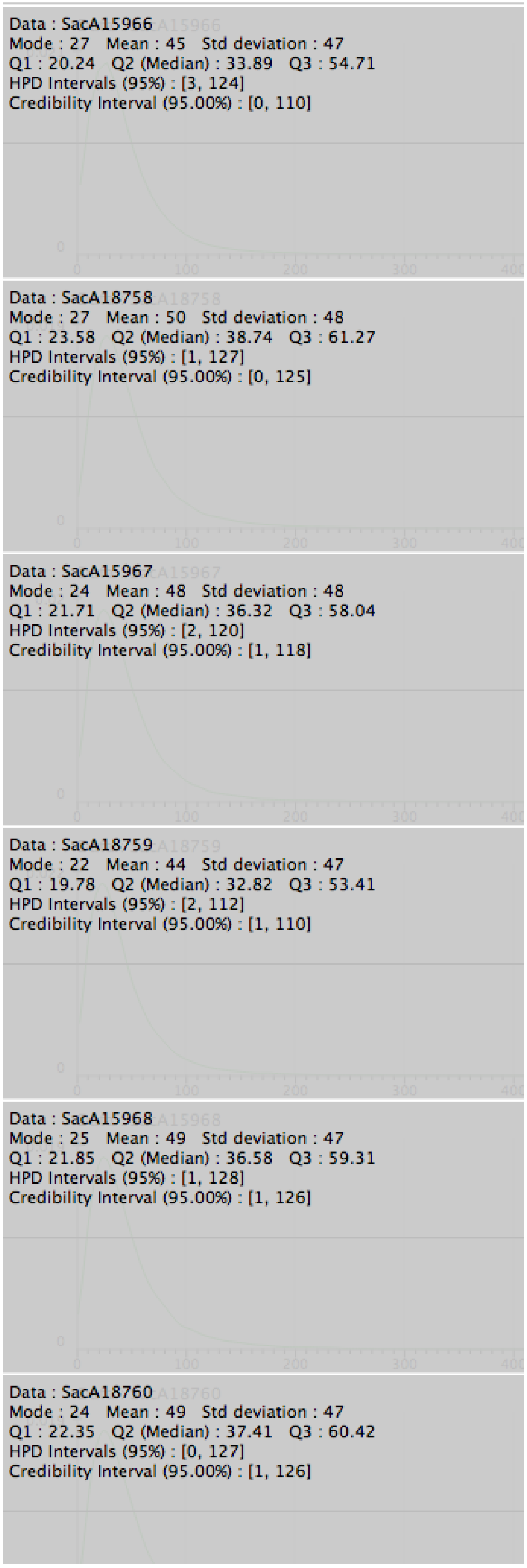


Figure 5.17 – Numerical values of indivudal standard deviations related to the modellisation of Bouquet 1

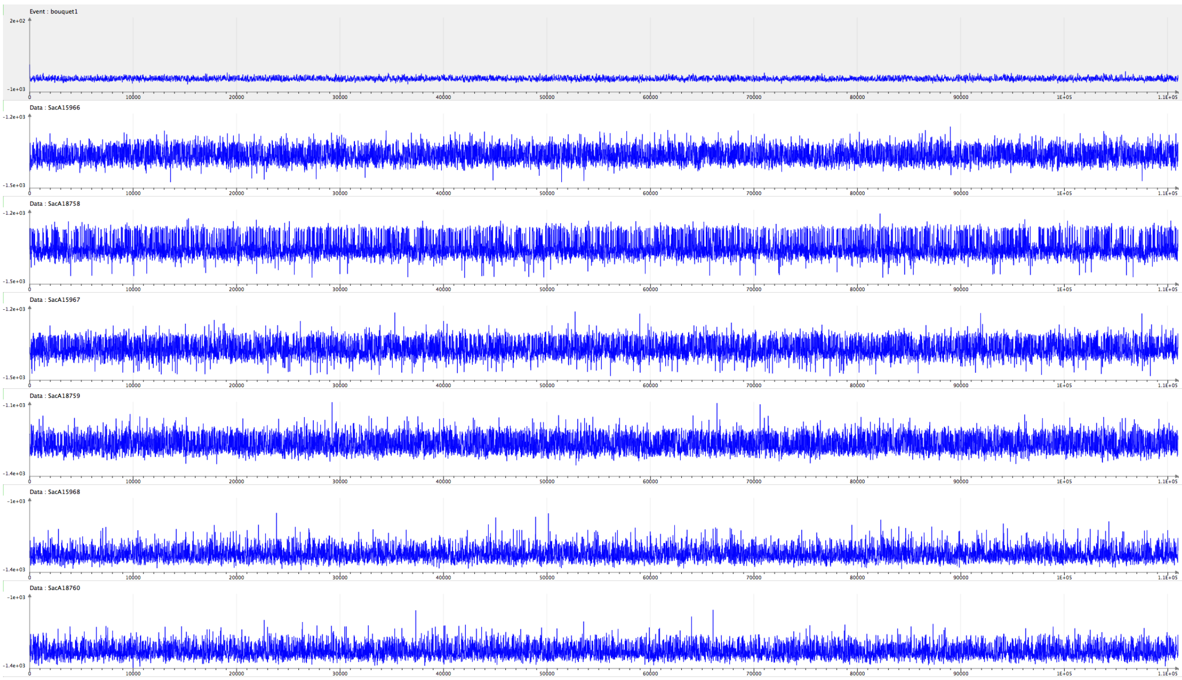


Figure 5.18 – History plots of indivudal standard deviations related to the modellisation of Bouquet 1

Hence, according to ChronoModel, all datations seem to be contemporary. Now we can draw conclusions about the calendar date of the event. The event Bouquet 1 may be dated at -1370 (mean value) with a 95% interval of [-1417; -1314] (HPD interval).

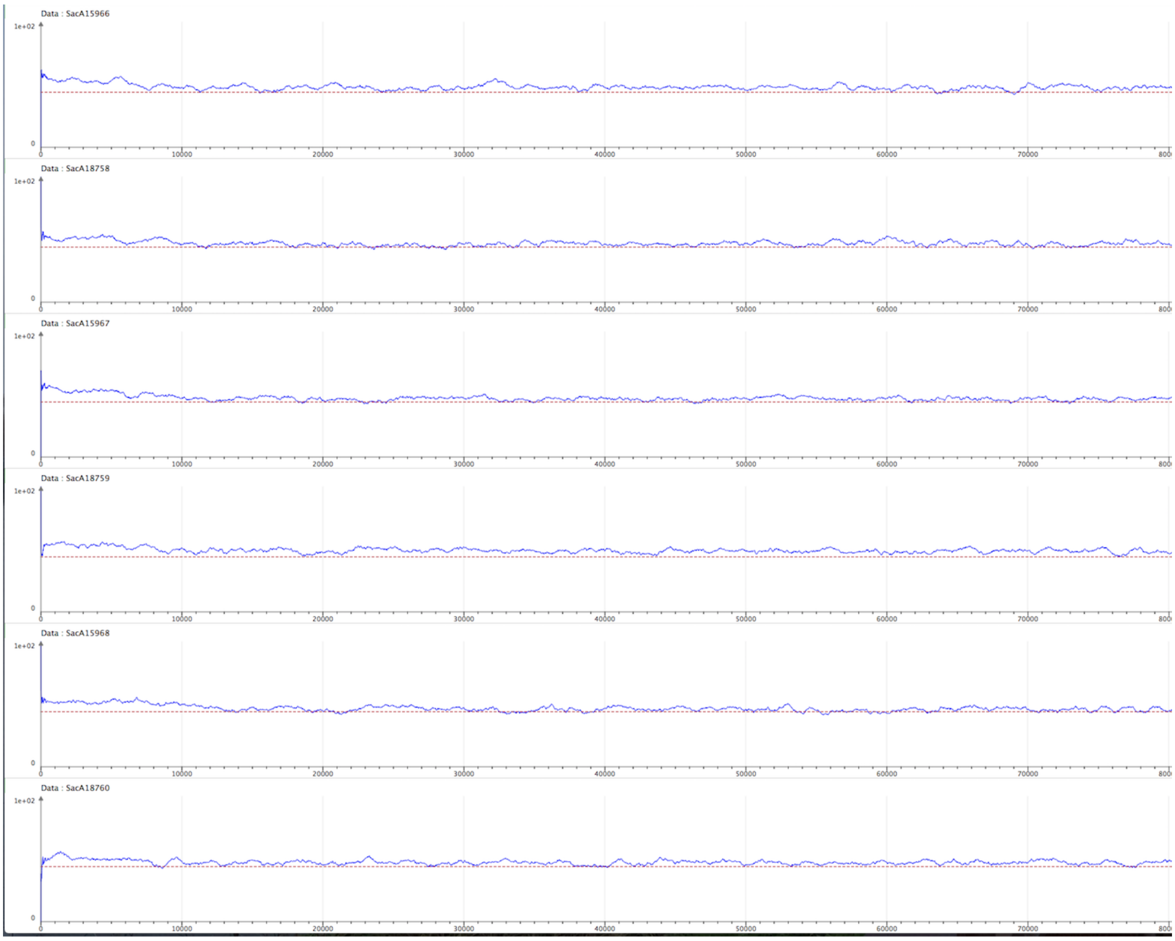


Figure 5.19 – Acceptation rates of indivudal standard deviations related to the modellisation of Bouquet 1

Figure 5.20 – Autocorrelation functions of indivudal standard deviations related to the modellisation of Bouquet 1

5.2.2 Bouquet 2

Now, let's say we want to estimate the calendar date of bouquet 2. 8 samples were extracted from Bouquet 2 and radiocarbon dated. The modelisation of this bouquet by ChronoModel is represented by Figure 5.21.

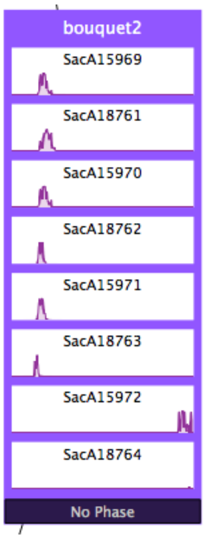


Figure 5.21 – Modelisation of Bouquet 2 with ChronoModel

Each radiocarbon measurement are calibrated using IntCal09 curve. No reservoir offset is taken into account. The study period is chosen to start at -2000 and end at 2000 using a step of 1 year. This study period is chosen so that every distribution of the calibrated date is included in this study period.

The method used to draw values from the conditional posterior distribution of the event Bouquet 1 is the default one, the rejection sampling using a double exponential proposal. The method used to draw values from the conditional posterior distribution of the datations is also the default one: Metropolis-Hastings algorithm using the posterior distribution of calibrated dates.

We use 1 000 iterations in the Burn-in period, 1 000 iterations in each of the 100 maximum batches in the Adapt period and 100 000 iterations in the Acquire period using thinning intervals of 1.

The marginal posterior densities, presented in Figure 5.22, are of two sorts. Althought, the first 6 datations seem to be contemporary, the two last ones seem to be some kind of outliers. Indeed their density function take values starting about 1600 whereas the other densities take values between -1500 and -1200. All history plots havegood mixing properties, acceptation rates are close to 43% and autocorrelation functions are correct (results not shown). Looking at individual standard deviations, the marginal posterior densities, displayed Figure 5.24, show three distinct standard

deviations. The first 5 samples are associated with a variance density function that takes small values, with mean values about 50. The next sample's individual variance has a mean posterior density at 100. And the 2 last samples are associated with individual standard deviations with a mean higher than 2 000 (See Figure 5.25 for numerical values). Hence these two last datations give a piece of information that has a reduced importance in the construction of the posterior density function of the event Bouquet 2. All individual standard deviations have a history plot with good properties, an acceptation rate about 43% or higher and a correct autocorrelation function (results not shown).
As a conclusion, the first 6 samples seem to be contemporary but the two last ones seem to be some kind of outliers. Then the event Bouquet 2 may be dated at -1344 (mean value) with a 95% HPD interval [-1405;-1278].

The example shows that the modelisation is robust to outliers. Indeed even if two outliers were included in the analysis, the datation of the event was not affected. ChronoModel do not need any particular manipulation of outliers before analysing the datations. Indeed there is no need to withdraw any datation or to use a special treatment to them (as done in Oxcal).

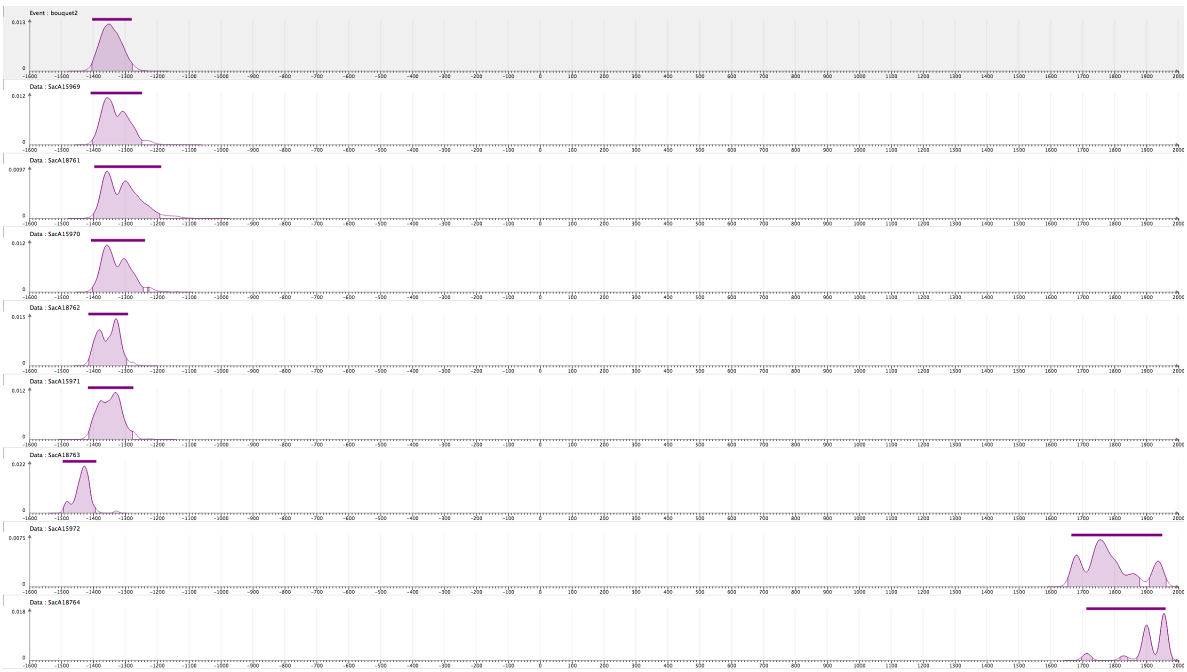


Figure 5.22 – Marginal posterior densities related to the modellisation of Bouquet 2

Figure 5.23 – Numerical values related to the modellisation of Bouquet 2

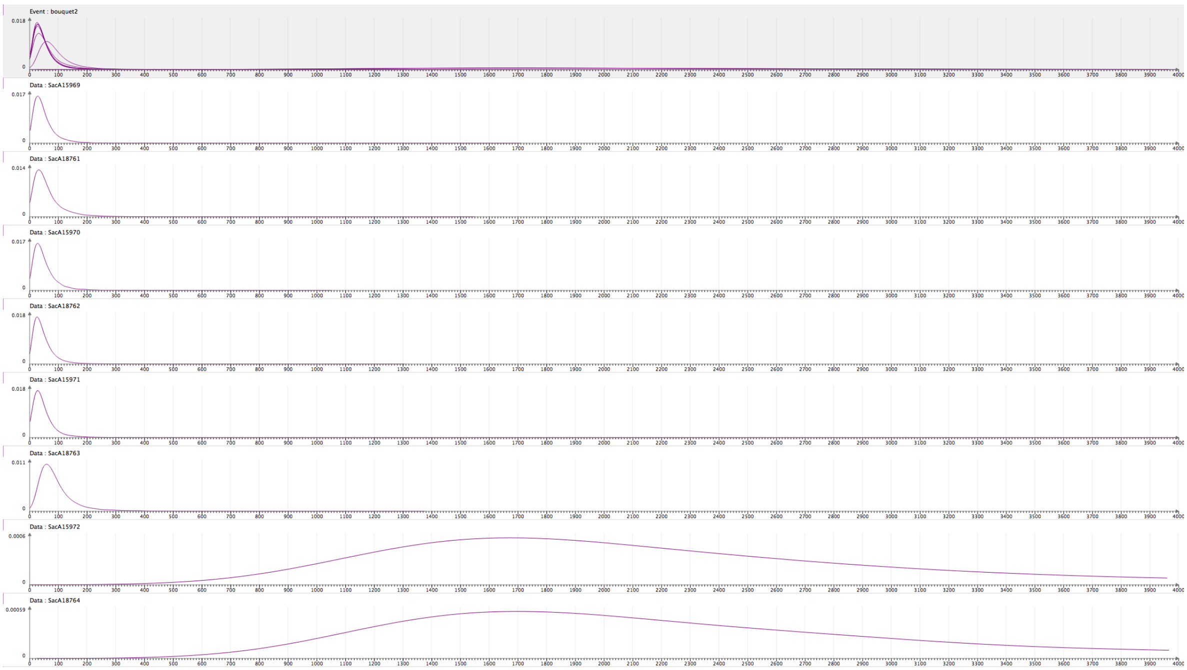


Figure 5.24 – Marginal posterior densities of indivudal standard deviations related to the modellisation of Bouquet 2

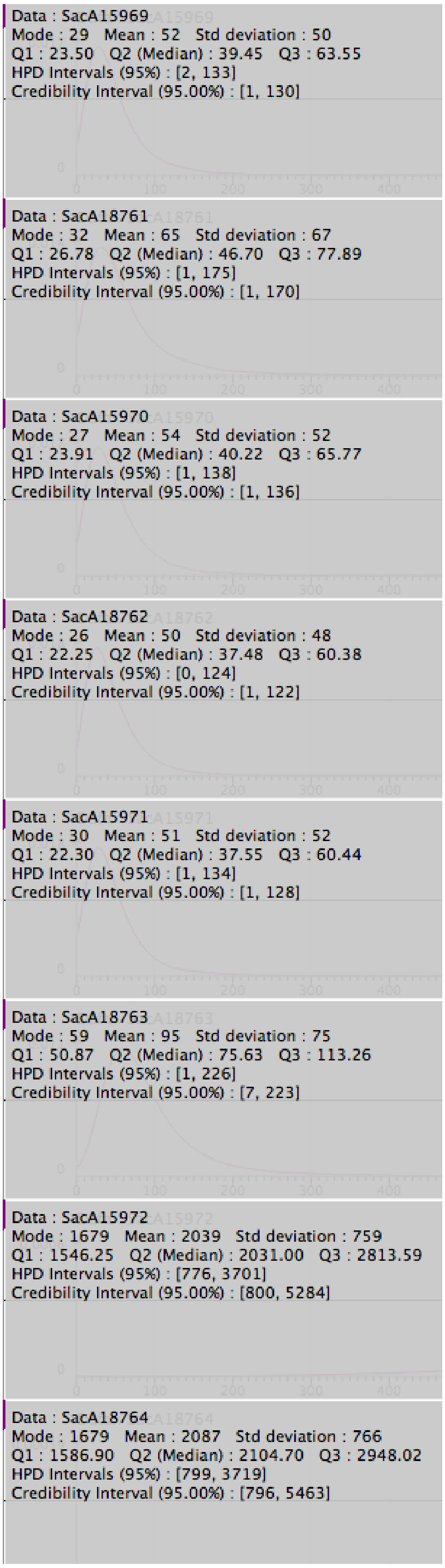


Figure 5.25 – Numerical values of indivudal standard deviations related to the modellisation of Bouquet 2

5.2.3 Modellisation of bouquets 1 and 2 simultaneously

Now let's say we want to date Bouquet 1 and Bouquet 2 simultaneously. See bouquets12.chr.
The study period was chosen to start at -2000 and end at 2000 using a step of 1 year. All other parameters are those used for the modellisation of Bouquet 1 and of Bouquet 2.

Three different modellisations are compared here. In the first modellisation, no further constraints are included (See Figure 5.26). In the two next modellisations, two bounds are introduced to constrain the beginnig and the end of the burial of Sennefer (See Figure 5.27). Indeed, the burial of Sennefer is assumed to have happened between the accession date of Tutankamun and the accession date of Horemheb (See [15]). These accesssion dates are concidered as bounds in ChronoModel. There are two different ways to introduce a bound. A bound may be a fixed date (Accession date of Tutankamun -1356 and Accession date of Horemheb -1312) or a bound may have a uniform distribution (Accession date of Tutankamun : uniform on [-1360; -1352], Accession date of Horemheb : uniform on [-1316;-1308]).

Figure **??** displays the marginal posterior densities of both bouquets when the modellisation does not include bounds. Figure **??** displays the marginal posterior densities of both bouquets when the modellisation does include bounds, using fixed bounds and using uniform bounds. From these results, we can see that the introduction of bounds helps restrain the posterior densities and the HPD interval of both events. However, using fixed bounds or bounds having a uniform distribution with a small period (8 years) lead to similar results. Numerical values are presented in Table 5.1.

	Modellisation of Bouquets 1 and 2		
	Without bounds	With fixed bounds	With uniform bounds
Event Bouquet 1			
Mean	-1371	-1338	-1338
HPD region	[-1418; -1315]	[-1357; -1316]	[-1358; -1316]
Event Bouquet 2			
Mean	-1344	-1336	-1336
HPD region	[-1406; -1278]	[-1356; -1313]	[-1357; -1313]

Table 5.1 – Numerical values related to the modellisations of Bouquets 1 and 2

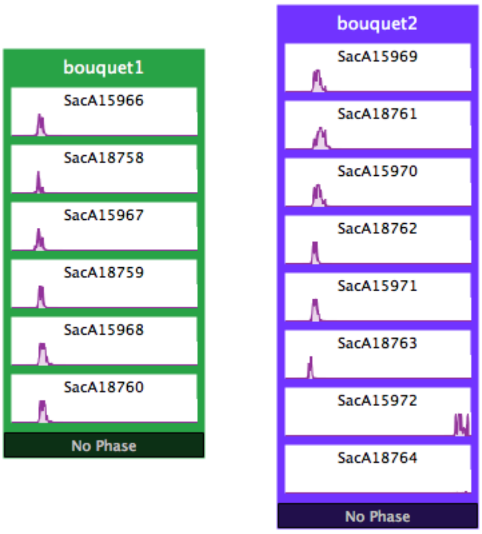


Figure 5.26 – Modelisation of Bouquets 1 and 2 without bounds

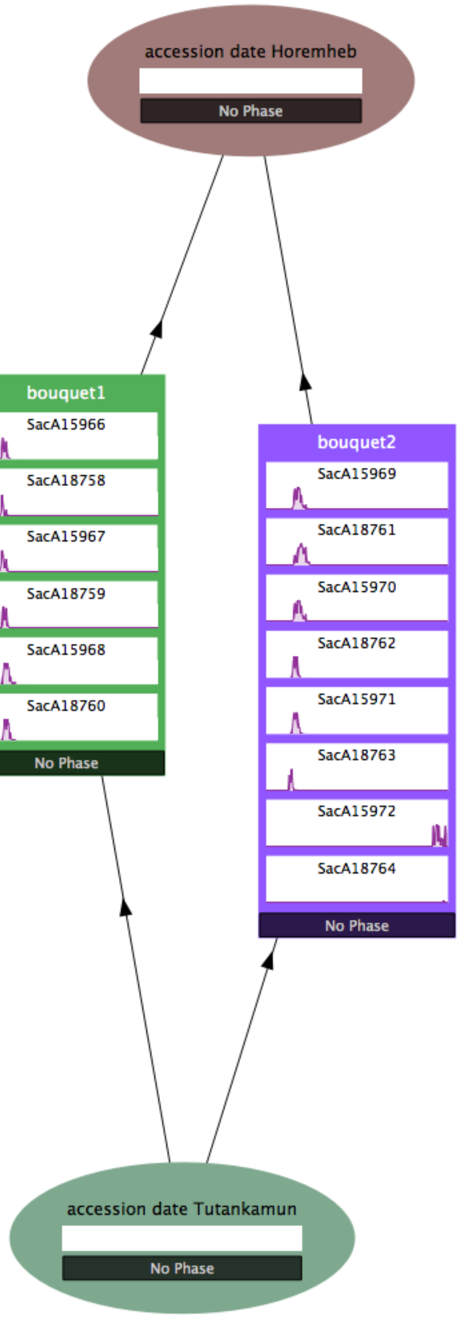


Figure 5.27 – Modelisation of Bouquets 1 and 2 including bounds

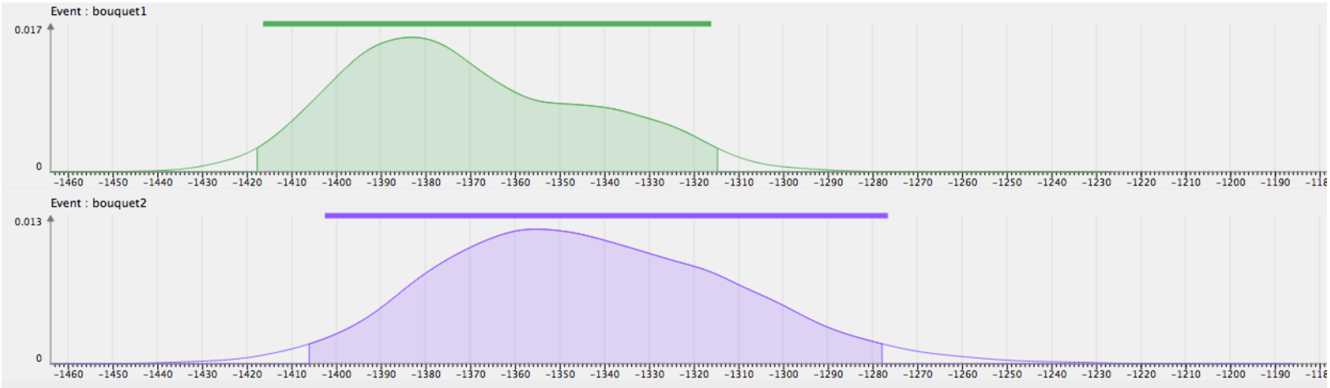


Figure 5.28 – Marginal posterior densities related to the modellisation of Bouquets 1 and 2 without bounds

Figure 5.29 – Marginal posterior densities related to the modellisation of Bouquets 1 and 2 with bounds (fixed bounds on the left handside figure, with uniform bounds on the right handside figure)

5.2.4 Modellisation of the phase including bouquets 1 and 2 - Estimation of the duration of the phase

Let's say that now we want to estimate the duration of the phase including both bouquets. This phase might be seen as the duration of Sennefer's burial. Three different modellisations are possible and presented here in turn. For all of them, the study period is chosen to start at -2000 and end at 2000 using a step of 1 year. All other parameters are those used for the modellisation of Bouquet 1 and of Bouquet 2.

5.2.4.1 Phase without constraints

The modellisation the phase including both Bouquets is displayed in Figure 5.30. In this modellisation, no further constraints are included. The phase's duration is kept unknown.

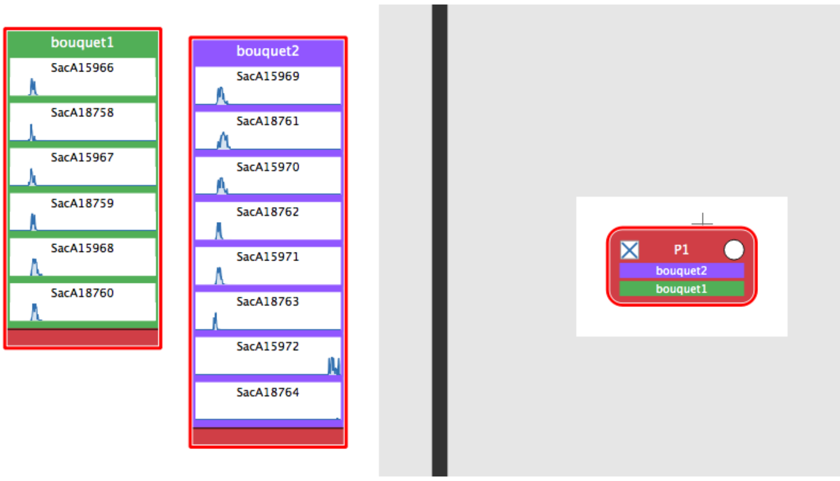


Figure 5.30 – Modelisation of the phase including Bouquets 1 and 2 including a phase

Figure 5.31 displays the marginal posterior densities of the both events and those of the beginning and the end of the phase. Statistical results regarding Bouquet 1 and Bouquet 2 are unchanged by the introduction of the phase, the results are similar to those presented in the last section when no bounds were introduced (See Table 5.1). In addition, this modellisation allows to estimate the mean duration of the phase (35) as well as its credibility interval ([0, 101]).

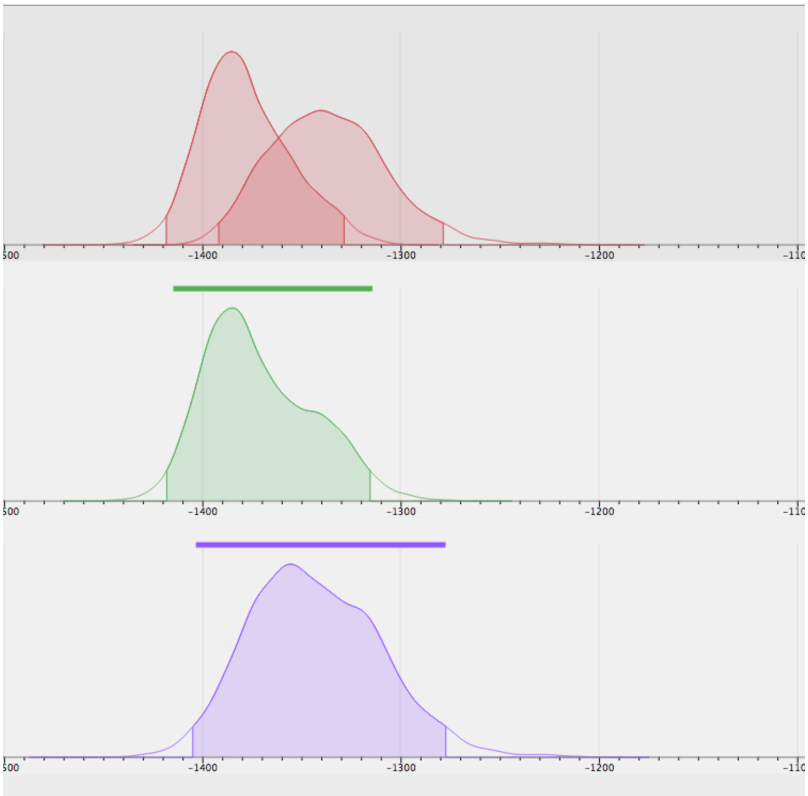


Figure 5.31 – Marginal posterior densities related to the modellisation of Bouquets 1 and 2 including a phase and without bounds. The densities of the minimum and the maximum are drawn in red, the density of Bouquet 1 is drawn in green, the density of Bouquet 2 is drawn in purple.

5.2.4.2 Phase with bounds

Now, let's include information about the accession dates of Tutankamun and Horemheb. We include the two fixed bounds as detailled in section 5.2.3. Here two modellisations are possible using ChronoModel. Figure 5.32 represents the first modellisation in which bounds constrain events. Figure 5.33 represents the second modellisation in which bounds are included in separated phases, using one phase for each bound, and stratigraphic constraints are placed between phases. However, these two modellisations give similar results. The marginal posterior densities of all parameters are presented in Figure 5.34. The mean duration of the bouquets' phase is 12 years associated with a credibility interval of [0, 33] that is smaller than the one estimated without bounds.

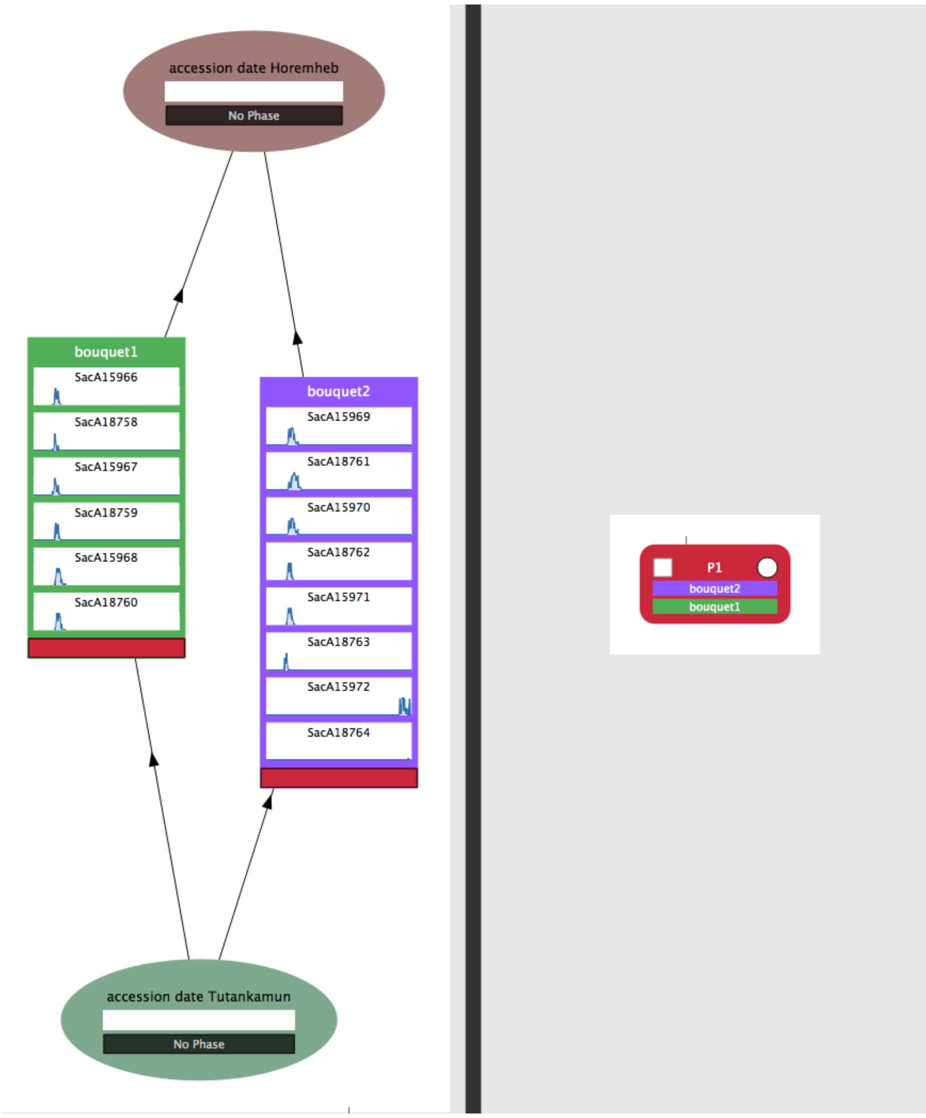


Figure 5.32 – First modelisation 1 of Bouquets 1 and 2 including a phase and bounds

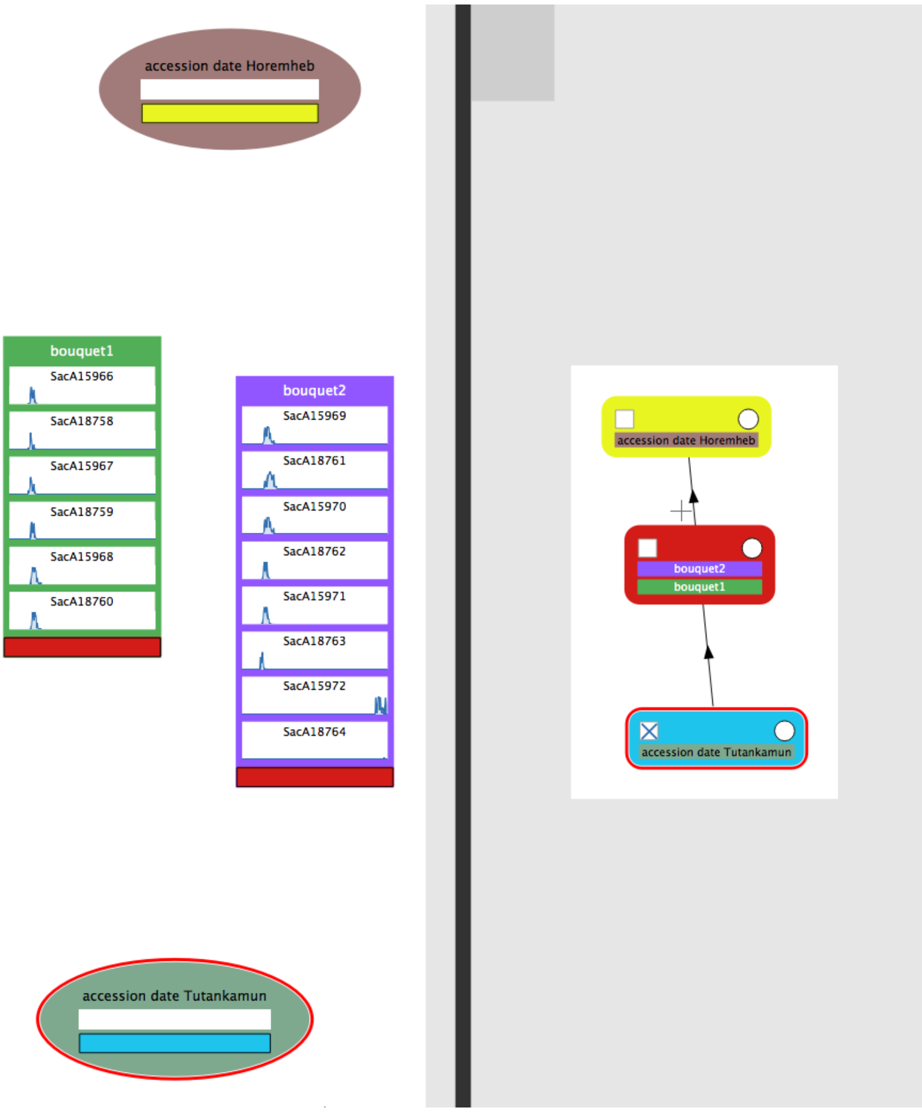


Figure 5.33 – Second modelisation of Bouquets 1 and 2 including a phase and bounds

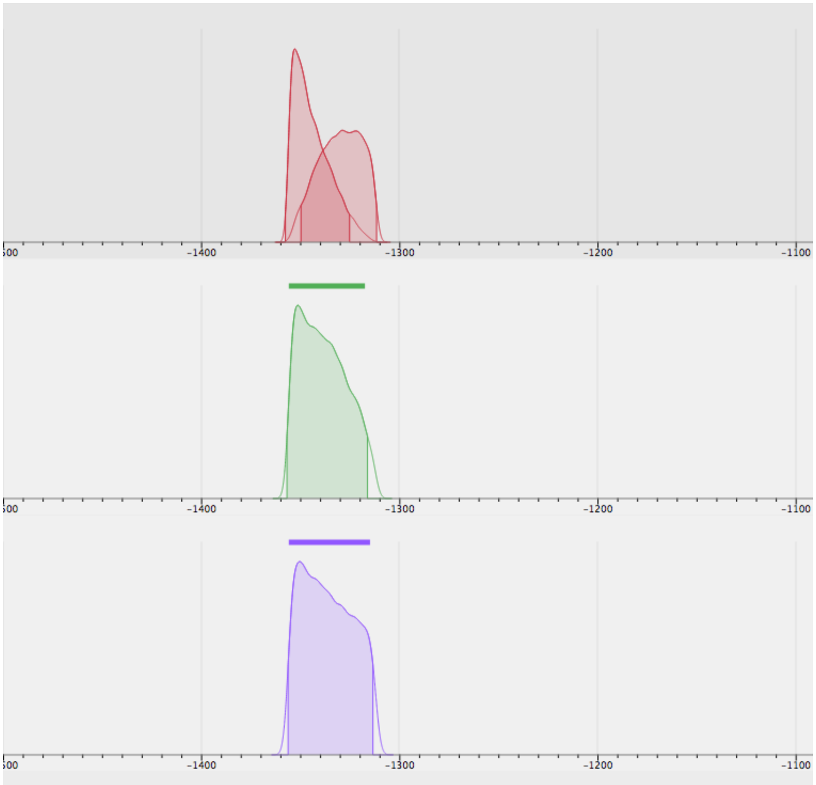


Figure 5.34 – Marginal posterior densities related to the modellisation of Bouquets 1 and 2 including a phase and bounds. The densities of the minimum and the maximum are drawn in red, the density of Bouquet 1 is drawn in green, the density of Bouquet 2 is drawn in purple.

5.2.4.3 Phase with fixed duration

The phase of Sennefer's burial is assumed to have happened between the accession date of Tutankamun (-1356) and the accession date of Horemheb (-1312). Hence, the duration of the phase is smaller than 44 years.
In this last modellisation, no bounds are included but the maximum duration of the phase is fixed at 44 years.

Figure 5.35 – Modelisation of the phase including Bouquets 1 and 2 having a fixed duration

This modellisation leads to the following results:

Duration of the phase		
Mean	20	
Credibility interval	[0; 41]	
Event	**Bouquet 1**	**Bouquet 2**
Mean	-1364	-1357
HPD region	[-1408; -1317]	[-1403; -1306]

Table 5.2 – Numerical values related to the phase including Bouquets 1 and 2 and having a fixed duration

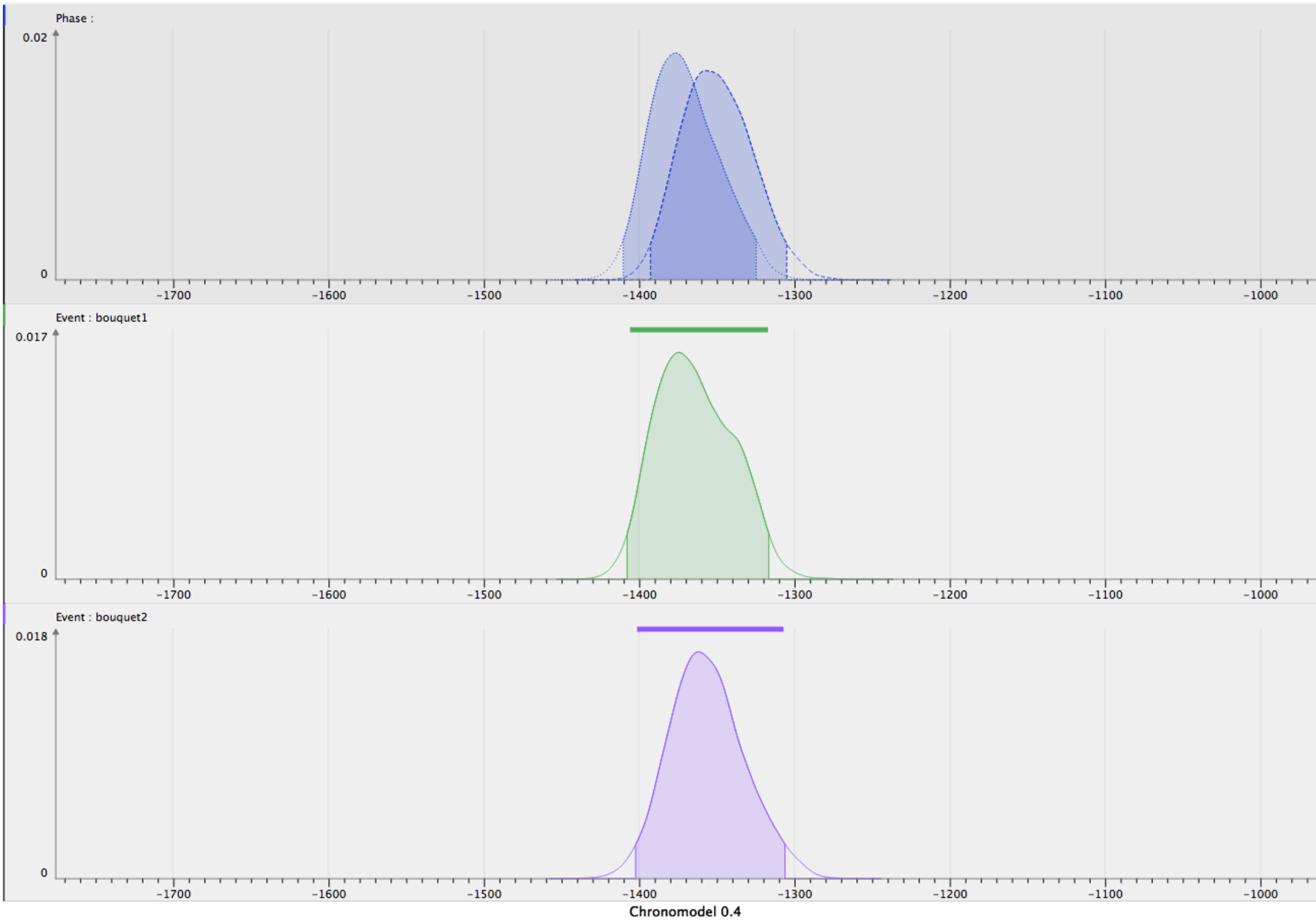
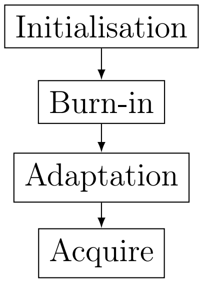


Figure 5.36 – Marginal posterior densities related to the modellisation of the phase including Bouquets 1 and 2 having a fixed duration.. The densities of the minimum and the maximum are drawn in red, the density of Bouquet 1 is drawn in green, the density of Bouquet 2 is drawn in purple.

Appendix A

Mathematical details related to MCMC algorithms

A.1 MCMC on variables



Algorithm 1 MCMC main sequence

```
variable initalisation
if Initialisation unsucced then
    Exit
for i ← 1, burn iteratins do                                        ▷ The burn-in loop
    Update all variable
repeat                                                          ▷ The adaptation loop
    for i ← 1, batch iteration do
        Update all variable
        Memory variables with adaptation corrections
        increment total iteration
    for all Variables do
        Compute acceptation rate
        if acceptation rate ≤ 0.44 OR acceptation rate ≥ 0.46 then
            Modify MH varialble
until all adaptation ≤ 0.46 OR total iteration = max batch
for i ← 1, iteration do                                             ▷ The acquire loop
    for i ← 1, thinning iteration do
        Update all variable
        increment total iteration
    for all Variables do
        Compute acceptation rate
        Memory variables
Show the results                                                    ▷ This is the end
```

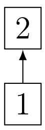
A.2 Vérification des contraintes sur les Faits et Bounds

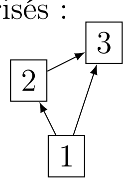
A.2.1 Relations stratigraphiques

Ces vérifications fonctionnent aussi bien entre les phases qu'entre les Faits. Le diagramme stratigraphique (contrainte d'ordre entre des Faits et/ou Bounds) ou de succession (contrainte d'ordre entre Phases) à la structure d'un DAG (Graphe acyclique orienté).
La relation d'ordre ($\leq$), doit être réflexive, antisymétrique et transitive, d'où pas de symétrie (a) ni de circularité (b)

A la construction du modèle, on peut tout de suite tester sur les faits et Bounds:

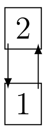
- Les cas autorisés :

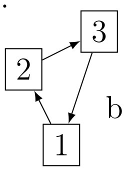




La relation de transitivité est inutile, car redondante sauf si introduction d'un hiatus γ

- Les cas interdits:

a




b

Enfin, il n'est pas possible de mettre un hiatus γ entre deux phases si un même Fait appartient à ces deux phases !

A.2.2 Vérifications sur les valeurs des Bounds

A.2.2.1 Contraintes stratigraphiques

Dans le cas de contraintes d'ordre sur les Bounds, il faut vérifier les conditions suivantes

- Si les Bounds θ^B sont des dates fixes, il est facile de vérifier si la contrainte strati entre θ_j^B et θ_k^B est possible.
  Il faut vérifier $\theta_j^B \leq \theta_k^B$ $(j \neq k)$
- Si les Bounds sont tels que: $\theta_{jm}^B \leq \theta_j^B \leq \theta_{jM}^B$

  Pour chaque θ_j^B dans une séquence (branche) strati, on doit vérifier:

$$\max_{i \leq j}(\theta_{im}^B) \leq \min_{k \geq j}(\theta_{kM}^B)$$

Si l'inégalité n'est pas vérifiée, alors envoi d'un message d'erreur signalant les Bounds concernés.

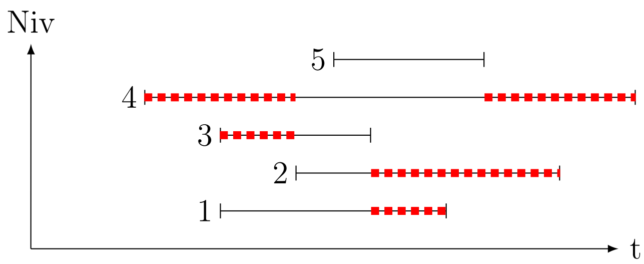
Une fois tout le modèle construit, on effectue un rétrécissement des intervalles des θ_j^B pour tous les j de toutes les séquences, soit :

$$\theta_{jm}^{Bnew} = \max_{i \leq j}(\theta_{im}^B)$$

$$\theta_{jM}^{Bnew} = \min_{k \geq j}(\theta_{kM}^B)$$

Ces nouvelles limites seront celles utilisées pour les vérifications sur les hiatus dans la section suivante.

Exemple:



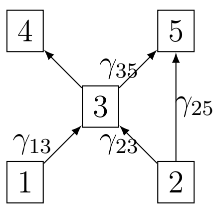
A.2.2.2 Hiatus

Des informations supplémentaires peuvent être apportées sur les contraintes d'ordre, à savoir que l'écart de temps γ minimal entre 2 phases (groupe de un ou plusieurs Faits et/ou Bounds), appelé hiatus, est:

- en succesion simple: les conditions suivantes seront vérifiées en posant $\gamma_m = \gamma_M = 0$
- connu fixé : $\gamma_m = \gamma_M = \gamma_0$
- incertain, uniforme entre 2 valeurs : $\gamma_m \leq \gamma \leq \gamma_M$

Les hiatus introduits entre phases doivent vérifier 2 conditions :

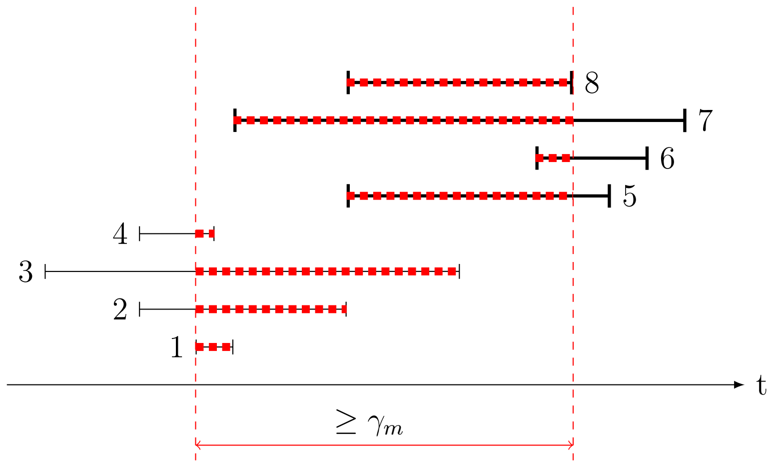
1. Il faut vérifier pour toutes les branches que (On vérifie sur la branche max): $\sum \gamma_m \leq (t_M - t_m)$ la période d'étude.



   Les branches identifiées sont 1-3-4; 1-3-5; 2-3-4; 2-3-5; 2-5

2. En présence de Bounds, une condition supplémentaire doit être vérifiée. Il faut s'assurer que l'écart γ_m entre 2 Bounds θ^{B1} et θ^{B2} de deux phases successives 1 et 2 soit possible.

   - si les Bounds θ^{B1} et θ^{B2} sont connus à l'année près (sans erreur), il faut vérifier:
   $$\min_j(\theta_j^{B2}) - \max_j(\theta_j^{B1}) \geq \gamma_m$$
   - Si les Bounds $\theta^B \in [\theta_m^B, \theta_M^B]$,



Les Bounds 1, 2, 3 et 4 sont dans la phase 1, les Bounds 5, 6, 7 et 8 sont dans la phase 2.
il faut vérifier :

$$\min_j(\theta_{jM}^{B2}) - \max_j(\theta_{jm}^{B1}) \geq \gamma_m$$

Si test Ok, on effectue un nouveau rétrécissement des intervalles des θ_j^B des deux phases, soit :

$$\theta_{jM}^{B1new} = \inf(\max_j(\theta_{jm}^{B1}), \theta_{jM}^{B1})$$

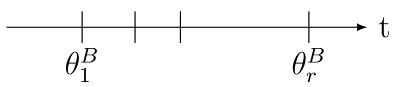
$$\theta_{jm}^{B2new} = \sup(\min_j(\theta_{jM}^{B2}), \theta_{jm}^{B2})$$

Ces nouvelles limites seront celles utilisées pour les vérifications sur les durées dans la section suivante.

A.2.2.3 Durées

Dans le cas de phases avec durée $\tau \in [\tau_m, \tau_M]$. Les Bounds doivent vérifier les conditions suivantes:

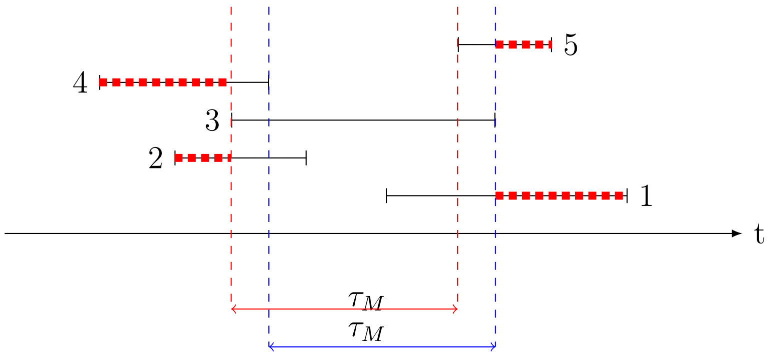
- Si les Bounds θ^B d'une phase sont fixés.



Il faut vérifier que :

$$\max_{j=1\ldots r}(\theta_j^B) - \min_{j=1\ldots r}(\theta_j^B) \leq \tau_M$$

- Si les Bounds θ^B sont tels que $\theta^B_m \leq \theta^B \leq \theta^B_M$



Il faut vérifier :

$$\max_{j=1\ldots r}(\theta^B_{jm}) - \min_{j=1\ldots r}(\theta^B_{jM}) \leq \tau_M$$

Si test Ok, on effectue un nouveau rétrécissement des intervalles $[\theta^B_{jm}, \theta^B_{jM}]$ en $[\theta^{Bnew}_{jm}, \theta^{Bnew}_{jM}]$ avec :

$$\theta^{Bnew}_{jm} = \sup(\max_j(\theta^B_{jm}) - \tau_M, \theta^B_{jm})$$

$$\theta^{Bnew}_{jM} = \inf(\min_j(\theta^B_{jM}) + \tau_M, \theta^B_{jM})$$

A.3 Variables initialisation

A.3.1 Main algorithm

A.3.2 Durée de phase

S'il y a une information sur la durée d'une phase,

- soit la durée est connue et fixée dans ce cas $\tau = \tau_{fix}$
- soit la durée est connue dans un intervalle donnée $\tau \in [\tau_{min}, \tau_{Max}]$. Nous autorisons le maximum de possibilité en posant $\tau = \tau_{Max}$

Listing A.1 – ChronoModel procedure to set the span

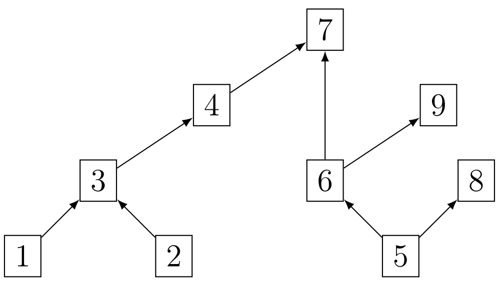
```
void Phase::initTau()
{
    if(mTauType == eTauUnknown)
    {
        // Nothing to do!
    }
    else if(mTauType == eTauFixed && mTauFixed != 0)
        mTau = mTauFixed;
    else if(mTauType == eTauRange && mTauMax > mTauMin)
        mTau = mTauMax;
}
```

A.3.3 Hiatus

S'il y a une information sur l'écart entre deux phases,

- soit l'écart est connue et fixée dans ce cas $\gamma = \gamma_{fix}$
- soit l'écart est connue dans un intervalle donnée $\gamma \in [\gamma_{min}, \gamma_{Max}]$. Nous autorisons le maximum de possibilité en posant $\gamma = \gamma_{min}$

A.3.4 Bounds



Il faut classer les Faits 1 à 9 dans l'ordre croissant des séquences strati en intercalant si nécessaire.Ici on peut identifier 5 séquences :

A 1, 3, 4, 7

B 2, 3, 4, 7

C 5, 6, 7

D 5, 6, 9

E 5, 8

On peut donc avoir le classement: 1, 2, 3, 5, 6, 4, 7, 9, 8 qui respecte l'ordre des séquences partielles.

1. Pour chaque Bound définir un niveau n, correspondant à un ordre d'initialisation. Nous pouvons prendre par exemple le niveau $n = 1$ pour le premier Bound à initialiser et augmenter le niveau au fur et à mesure de l'ordre dans la séquence.

2. Pour les Bounds de niveau 1,(n ° 1, 2, 5 dans notre figure)

$$\theta_1^B = Unif[\theta_{1m}^B, \theta_{1M}^B]$$

   Pour les autres Bounds $n \geq 2$(après réduction d'intervalle) :

$$\theta_n^B = Unif[\max(\theta_{n-1}^B, \theta_{nm}^B), \theta_{nM}^B]$$

Algorithm 2 Initializing θ of Bounds- Part 1

```
                                      ▷ 1- We reduce the Bound ascending levels
for P ← 1, ProfMax do
    for all Bound with Prof = P do
        reduce the Bound
                                      ▷ 2- We reduce the Bound owned by phases
for all Phase with Bound as element do
    if τ^P is known then
        Bi ← find the Maximun of mininimal values of bounds in this phase
        Bx ← find the Minimun of maxinimal values of bounds in this phase
        if Bi − Bx ≥ τ_Max then
            init impossible with phase
            if BLi ≤ Bi − τ_Max then ▷ Reduice the Bound with span bigger than τ_Max
                BLi = Bi − τ_Max
            if BLx ≤ Bx + τ_Max then
                BLx = Bx + τ_Max
            BPi ← BLi
            BPx ← BLx
                                      ▷ 3- We define the Bound ascending levels
for P ← 1, ProfMax do
    for all Bound with Prof = P do
        θ ← find the max of θ of Bound in the superior Constraints ▷ If Bound belongs to a phase in constraint with other phases, we look also into the next phases
        for all Phase with this Bound as element do
                                      ▷ Bound can belong to several Phase
```

Algorithm 3 Initializing θ of Bounds - Part 2

$\triangleright$ 3.1-Here we have to find the constraint due to the τ^P of Phase

if τ^P *is known* **then**

$\alpha \leftarrow$ *the minimal* θ*in this phase*

$\beta \leftarrow$ *the maximal* θ*in this phase*

$\triangleright$ Save the extremum values

if $\theta_{max}^{Phase} > \alpha + \tau^P$ **then**

$\theta_{max}^{Phase} \leftarrow \alpha + \tau^P$

if $\theta_{min}^{Phase} < \beta - \tau^P$ **then**

$\theta_{min}^{Phase} \leftarrow \beta - \tau^P$

$\triangleright$ 3-2 tirage du θ limite entre le inf(θ en dessous) et la Borne Bmax réduite $\triangleright$ 3.2 comparaison and sampling θ value

if $\theta_{min}^{Phase} \leq \theta_{inf}$ **then**

$\theta_{inf} \leftarrow \theta_{min}^{Phase}$

if $\theta_{max}^{Phase} \geq \theta_{sup}$ **then**

$\theta_{sup} \leftarrow \theta_{max}^{Phase}$

$\theta = Unif[\theta_{inf}; \theta_{sup}]$

A.3.5 Faits

1. Si aucun fait initialisé, alors on pose :

$$\theta_{inf} = t_m \text{ et } \theta_{sup} = t_M$$

   sinon on calcule:

$$\theta_{inf} = \sup[\max_{strati}(\theta_{prec}), \max_{phases}(\theta^P_{k\neq j}) - \tau^P, \max_{\text{phases précedantes}}(\theta^{P-1}) + \gamma^{p-1}, t_m]$$

   et

$$\theta_{sup} = inf[\min_{strati}(\theta_{suiv}), \min_{phases}(\theta^P_{k\neq j}) + \tau^P, \min_{\text{phases suivantes}}(\theta^{P+1}) - \gamma^P, t_M]$$

   Formules qui ne s'appliquent que sur les faits ou Bounds déjà initialisés, sinon $\theta_{inf} = t_m$ et/ou $\theta_{sup} = t_M$.

2. Tirer θ_j^P tel que :

$$\theta_j^P = Unif[\theta_{inf}, \theta_{sup}]$$

   L'initialisation des faits peut se faire dans n'importe quel ordre.

Algorithm 4 Initializing θ of event

for all θ_{Event} **do**
$\triangleright$ 1 - Here we have to find the constraint due to the chronologie on Event, including Phase constraint

$\theta_{sup} \leftarrow$ findContrainteSup$(Event)$ $\triangleright$ If Event belongs to a phase in constraint with other phases findContrainteSup(l)ook also into the next phases
$\theta_{inf} \leftarrow$ findContrainteInf$(Event)$
$\theta^{Phase}_{min} \leftarrow t_M$ $\triangleright$ t_M: the maximun date of study
$\theta^{Phase}_{max} \leftarrow t_m$ $\triangleright$ t_m: the minimun date of study
for all Phase with θ_{Event} as element **do** $\triangleright$ θ_{Event} can belong to several Phases
$\triangleright$ 2 - Here we have to find the constraint due to the τ^P of Phase
if τ^P is known **then**
$\triangleright$ Temporary storage of the extreme values,we don't care about unintialized θ
$\alpha =$ minimalThetaInPhase$(Phase)$
$\beta =$ maximalThetaInPhase$(Phase)$
if $\theta^{Phase}_{max} > \alpha + \tau^P$ **then**
$\theta^{Phase}_{max} \leftarrow \alpha + \tau^P$
if $\theta^{Phase}_{min} < \beta - \tau^P$ **then**
$\theta^{Phase}_{min} \leftarrow \beta - \tau^P$
if $\theta^{Phase}_{min} \leq \theta_{inf}$ **then** $\triangleright$ 3 - Comparing θ_{inf} found
$\theta_{inf} \leftarrow \theta^{Phase}_{min}$
if $\theta^{Phase}_{max} \geq \theta_{sup}$ **then**
$\theta_{sup} \leftarrow \theta^{Phase}_{max}$
$\theta = Unif[\theta_{inf}; \theta_{sup}]$ $\triangleright$ 4 - Sampling θ value

A.3.6 Date

Initialisation de la date t_i par tirage MH dans la densité calibrée

A.4 Algorithme MCMC

A.4.1 Algorithme principal de mise à jour

A.4.2 Durée de phase

On a $0 \leq \tau_m^P \leq \tau_M^P \leq (t_M - t_m)$
Mais le support $[\tau_m^P, \tau_M^P]$ est restreint en fonction des θ^P:
On doit avoir $\tau^p \geq \max(\theta^P) - \min(\theta^P)$, d'où:

$$\tau_{inf} = \sup[\tau_m^P, \max(\theta^P) - Min(\theta^P)] \leq \tau_{new}^P \leq \tau_M^P$$

$$\tau \sim Unif[\tau_{inf}, \tau_M^P]$$

Algorithm 5 Sampling τ

if τ^P unknown **then** ▷ Nothing sampling
else
  $\tau_{inf} \leftarrow \max(\theta^P) - \min(\theta^P)$;
  if $\tau_{inf} \leq \tau_m^P$ **then**
    $\tau_{inf} \leftarrow \tau_m^P$
  $\tau_{new}^P \leftarrow Unif[\tau_{inf}, \tau_M^P]$ ▷ Here the sampling

A.4.3 Hiatus

Le tirage ne dépend que de γ_m^P, γ_M^P et de l'état des faits θ des deux phases encadrantes P et P+1, d'où:

$$\gamma_m^P \leq \gamma_{new}^P \leq \inf[\gamma_M^P, \min(\theta^{P+1}) - \max(\theta^P)]$$

Algorithm 6 Sampling γ

if γ unknown **then** ▷ Nothing sampling
else if γ fixed **then**
  $\gamma \leftarrow \gamma$ ▷ not change the value
else if γ in a range $[\gamma_m; \gamma_M]$ **then**
  $\gamma_{sup} \leftarrow \min[\gamma_M; \min(\theta^{P+1}) - \max(\theta^P)]$; ▷ $\min(\theta^{P+1}) - \max(\theta^P)]$ always lessan γ_m
  $\gamma_{new} \leftarrow Unif[\gamma_m, \gamma_{sup}]$ ▷ Here the sampling

A.4.4 Bound

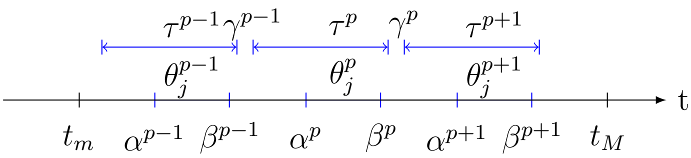
The update of Bound θ_j^{BP} must check the following constraints :

$$\sup[\max_{strati}(\theta_{prec}), \max_{phases}(\theta_{k\neq j}^P) - \tau^P, \max_{\text{phases précedentes}}(\theta^{P-1}) + \gamma^{P-1}, \theta_m^{BP}, t_m] \leq \theta_j^{BP}$$

et

$$\theta_j^{BP} \leq \ inf[\min_{strati}(\theta_{suiv}), \min_{phases}(\theta_{k\neq j}^P) + \tau^P, \min_{\text{phases suivantes}}(\theta^{P+1}) - \gamma^P, \theta_M^{BP}, t_M]$$

A.4.5 Fait



Les Faits doivent respecter les contraintes de support suivantes::

1. $\theta_j \in [t_m, t_M]$ plage d'étude.
2. les θ_j respectent les contraintes strati (ordre total ou partiel) $\max(\theta_{prec}) \leq \theta_j \leq \min(\theta_{suiv})$
3. les θ_j respectent les contraintes de durée τ: $\max(\theta^P) - \min(\theta^P) \leq \tau^P$
4. les θ_j respectent les contraintes de hiatus γ: $\max(\theta^{P+1}) - \min(\theta^P) \geq \gamma^p$

La mise à jour d'un Fait θ_j^P doit vérifier les contraintes suivantes:

$$\sup[\max_{strati}(\theta_{prec}), \max_{phases}(\theta_{k\neq j}^P) - \tau^P, \max_{\text{phases précedentes}}(\theta^{P-1}) + \gamma^{P-1}, t_m] \leq \theta_j^P$$

et

$$\theta_j^P \leq \; inf[\min_{strati}(\theta_{suiv}), \min_{phases}(\theta_{k\neq j}^P) + \tau^P, \min_{\text{phases suivantes}}(\theta^{P+1}) - \gamma^P, t_M]$$

A.4.6 Date

A.4.7 Décalage Wiggle Matching

A.4.8 Variance individuelle

A.4.9 Requête sur les phases

Une fois tous les θ_j mis à jour. On détermine le début, la fin et la durée. Dans ChronoModel, toutes les requêtes sont exécutées dans la même procédure. updateTau n'est pas un requête mais un nouveau tirage de τ (voir A.4.2).

A.4.9.1 Début

Le début d'une phase correspond au θ le plus petit des Faits inclus dans une phase. C'est à dire le première Fait observé chronologiquement dans une phase.

$$\hat{\alpha}^P = \min(\theta^P_{j,j=1...r})$$

A.4.9.2 Fin

La fin d'une phase correspond au θ le plus grand des Faits inclus dans une phase. C'est à dire le dernier Fait observé chronologiquement dans une phase.

$$\hat{\beta}^P = \max(\theta^P_{j,j=1...r})$$

A.4.9.3 Durée

La durée correspond au temps maximal séparant les faits d'une même phase.

$$\hat{\tau^P} = \hat{\beta^P} - \hat{\alpha^P}$$

Algorithm 7 Request on τ

if τ^P *unknown* **then**
    $\hat{\tau}^P \leftarrow \max(\theta^P) - \min(\theta^P)$ ▷ The request
else
    $\hat{\tau}^P \leftarrow \tau^P_{new}$ ▷ The result of the sampling

A.4.9.4 Hiatus

La durée correspond au temps minimal séparant les faits de deux phases successives.Le hiatus $\hat{\gamma} = \hat{\alpha}^{P+1} - \hat{\beta}^P$

Algorithm 8 Request on γ

```
if γ unknown then
    γ̂ ← max(θ^P) − min(θ^{P+1})          ▷ The request
else
    γ̂ ← γ_new                            ▷ The result of the sampling
```

Bibliography

[1] Christen JA. Summarizing a set of radiocarbon determinations: a robust approach. Applied Statistics. 1994;43(3):489–503.

[2] Lunn, David; Jackson, Chris ; Spiegelhalter D. The BUGS Book: A Practical Introduction to Bayesian Analysis; 2012.

[3] Daniels MJ. A prior for the variance in hierarchical models. Canad J Statist. 1999;27(3):567–578.

[4] Spiegelhalter DJ, Abrams KR, Myles JP. Bayesian Approaches to Clinical Trials and Health-Care Evaluation. Chichester: Wiley; 2004.

[5] Reimer P, Baillie M, Bard E, Bayliss A, Beck J, Bertrand C, et al. IntCal04 terrestrial radiocarbon age calibration, 0-26 cal kyr BP. Radiocarbon. 2004;46(3).

[6] Hughen K, Baillie M, Bard E, Beck J, Bertrand C, Blackwell P, et al. Marine04 marine radiocarbon age calibration, 0-26 cal kyr BP. Radiocarbon. 2004;46(3).

[7] McCormac FG, Hogg A, Blackwell P, Buck C, Higham T, Reimer P. SHCal04 Southern Hemisphere calibration, 0-11.0 cal kyr BP. Radiocarbon. 2004;46(3).

[8] Reimer P, Baillie M, Bard E, Bayliss A, Beck J, Blackwell P, et al. IntCal09 and Marine09 Radiocarbon Age Calibration Curves, 0-50,000 Years cal BP. Radiocarbon. 2009;51(4).

[9] Reimer P, Bard E, Bayliss A, Beck J, Blackwell P, Ramsey CB, et al. IntCal13 and Marine13 Radiocarbon Age Calibration Curves 0–50,000 Years cal BP. Radiocarbon. 2013;55(4).

[10] Hogg A, Hua Q, Blackwell P, Niu M, Buck C, Guilderson T, et al. SHCal13 Southern Hemisphere Calibration, 0–50,000 Years cal BP. Radiocarbon. 2013;55(4).

[11] Buck CE, Litton CD, G CW. The Bayesian Approach to Interpreting Archaeological Data. England: Chichester, J.Wiley and Son; 1996.

[12] Robert CP, Casella G. Monte Carlo Statistical Methods. 2nd ed. Springer; 2004.

[13] Robert CP. Simulation of truncated normal variables. Statistics and Computing. 1995;5(2):121–125.

[14] Roberts GO, Rosenthal JS. Optimal Scaling for Various Metropolis-Hastings Algorithms. Statistical Science. 2001;16(4):351–367.

[15] Quiles A, Aubourg E, Berthier B, Delque-Količ E, Pierrat-Bonnefois G, Dee MW, et al. Bayesian modelling of an absolute chronology for Egypt's 18th Dynasty by astrophysical and radiocarbon methods. Journal of Archaeological Science. 2013;40:423–432.