HAL
open science

Analysis of Archaeological Phases using the CRAN Package RChronoModel

Anne Philippe, Marie-Anne Vibet

To cite this version:

Anne Philippe, Marie-Anne Vibet. Analysis of Archaeological Phases using the CRAN Package RChronoModel. 2017. hal-01347895v2

HAL Id: hal-01347895

https://hal.science/hal-01347895v2

Preprint submitted on 12 Jan 2017 (v2), last revised 20 Sep 2018 (v5)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Analysis of Archaeological Phases using the CRAN Package RChronoModel

Anne Philippe
Université de Nantes, LMJL

Marie-Anne Vibet
Université de Nantes, LMJL

Abstract

We propose new statistical tools to analyze and to estimate archeological phases from the posterior distribution of a sequence of dates. All the functions are implemented in the R package **RChronoModel**. The required inputs are simulated samples from the posterior distribution of the dates. Such MCMC samples are provided, for instance, by ChronoModel or Oxcal, two softwares build for the chronological modeling of archeological dates. We give a practical introduction to the package **RChronoModel** using published data and comment the statistical results.

Keywords: Bayesian modeling, Gap interval, Gap test, MCMC samples, Tempo plot, Time range, Transition interval.

1. Introduction

The statistical modeling within the Bayesian framework is widely used by archaeologists for constructing chronologies. This consists in the estimation of a sequence of dates where

- the observations are the measurements coming from possibly different dating methods,
- the prior information comes from geological, historical, environmental or any other considerations.

Bayesian inference is a probabilistic estimation method, in the sense that the information on the estimated parameters is given by a probability distribution called posterior probability. Usually the analytic expression of the posterior distribution is not easily computable. However, inferences may be drawn using a sample from the posterior density. More precisely, it is possible to compute a Markov chain whose target distribution is the posterior distribution.

Indeed, Markov chain Monte Carlo (MCMC) algorithms provide a way of drawing samples from the joint posterior distribution in high-dimensional Bayesian models.

Two main softwares are available to estimate chronologies: Chronomodel (see Lanos, Philippe, Lanos, and Dufresne (2016); Vibet, Philippe, Lanos, and Dufresne (2016)) and Oxcal (see Ramsey (2016, 2009)). Both softwares return the marginal posterior distribution of all the estimated dates with elementary statistics (mean, standard deviation HPD region,...). It is also possible to export the output from MCMC simulations (see section 4 for the details). Thus, we can explore the joint distribution of the sequence of dates. For example we can construct:

- **Tempo plots** as defined by Thomas Dye (See Dye (2016)), that is "a statistical graphic designed for the archaeological study of rhythms of the long term that embodies a theory of archaeological evidence for the occurrence of events" and **their associated activity plots**.
- Time intervals to characterize archaeological phases or periods, called **time range intervals**.
- Time intervals to characterize the transition between two successive phases or periods, called **transition intervals**.
- **Testing procedures to check the presence of a gap** between two successive phases or periods. A **gap interval** is estimated if we accept its existence .

The CRAN package **RChronoModel** provides a list of functions for calculating these statistics from the Markov Chains simulated by Chronomodel or by Oxcal.

In Section 2, we propose new statistical tools for analysing some groups of estimated dates in a Bayesian framework. Section 3 provides a short description of the dataset used to illustrate **RChronoModel**. The data was published by Bosch, Mannino, Prendergast, O'Connell, Demarchi, Taylor, Niven, van der Plicht, and Hublin (2015). Section 4 specifies how the MCMC samples can be extracted from Chronomodel or by Oxcal. In Section 5, we give a practical introduction to the use of **RChronoModel** functions (version 0.4, published in January 2017).

For non-users of R software, a web application, ArchaeoPhases, has also been developed in order to take advantages of these functions without having to know R. This application is freely available on `https://archaeology-bayesian-modelling.shinyapps.io/ArchaeoPhases`

2. Statistical aspects

We define a chronology as a sequence of calendar dates $\tau_1, \ldots, \tau_n$. The archeological question about a chronology is not only a matter of estimation of calendar dates $\tau_1, \ldots, \tau_n$ but also the characterization of historical/geological periods called archaeological phases.

The estimation of the phase can only be done from the estimated dates associated to this phase.
In the following, we define a phase (denoted P) by a collection of dates, τ_i for $i \in I$ where I is a subset of $\{1, \ldots, n\}$ and n the total number of calendar dates.

A phase is commonly summarized by two parameters: its start and its end. Two different approaches are considered in order to estimate these two dates.

1. The start and the end are modeled, and so additional parameters are incorporated in the model. This is the approach implemented in Oxcal where the parameters are denoted t_a and t_b. This requires to do prior assumption on the distribution of the dates on the estimated period $[t_a, t_b]$.

2. The start (resp. the end) is estimated by the minimum (resp. the maximum) of the dates included in the phase. Thus the approach does not integrate a model of the phase, but it is only a post-processing step based on the estimation of a sequence of dates. These statistics are implemented in ChronoModel and Oxcal.

Our approach is in the same spirit of the second one in the sense that we propose statistical tools of post-processing without additional modeling.

Notation : we denote by $\mathcal{M}$ the set of measurements coming from dating methods. We assume that MCMC samples from the joint posterior distribution $p(\tau_1, \ldots, \tau_n \,|\, \mathcal{M})$ of all the dates $\tau_1, \ldots, \tau_n$ are available (using for instance ChronoModel or Oxcal).

2.1. Time Range interval of a phase

We propose to characterize a phase using a time interval called a **time range interval**. A time range interval summarizes the dates defining the phase by a time period having fixed posterior probability. It gives also an idea of the start, the end and the duration of a phase.

Definition 1. *Let $P = \{\tau_i \; i \in I \subset \{1, \ldots, n\}\}$ be a phase. The $100(1-\gamma)\%$ phase time range is the shortest interval $[a, b]$ such that*

$$P(\tau_i \in]a, b[, \; \forall i \in I) | \, \mathcal{M}) = 1 - \gamma \tag{1}$$

This means that the time interval $[a, b]$ contains all the dates $\{\tau_i \; i \in I\}$ with a fixed posterior probability $1 - \gamma$ (e.g. 95 %, 68 %...).

We denote $\alpha = \min(\tau_i, \forall i \in I)$ and $\beta = \max(\tau_i, \forall i \in I)$ The equation (1) can be also rewritten of the form

$$P(a \leq \alpha \leq \beta \leq b \,|\, \mathcal{M}) = 1 - \gamma \tag{2}$$

Therefore the construction of the time range interval depends only on the joint posterior distribution of (α, β).

The collection of intervals $[a, b]$ satisfying (2) is described by this set of intervals:

$$\{[a(\epsilon) \, ; \, b(\epsilon)], \quad \epsilon \in [0, \gamma]\}$$

where

- $a(\epsilon) = F_\alpha^-(\epsilon \,|\, \mathcal{M})$ with $F_\alpha^-(\cdot \,|\, \mathcal{M})$ is the quantile function of the posterior distribution of α

- $b(\epsilon) = F_\beta^-(\frac{1-\gamma}{1-\epsilon} \,|\, \alpha \geq a(\epsilon), \mathcal{M})$ with $F_\beta^-(\cdot \,|\, \alpha \geq a(\epsilon), \mathcal{M})$ is the quantile function of the conditional distribution of β given $\alpha \geq a(\epsilon)$ and $\mathcal{M}$.

Indeed, $\alpha \leq \beta$ happens almost surely, so we have

$$\begin{aligned} P(a(\epsilon) \leq \alpha \leq \beta \leq b(\epsilon) \,|\, \mathcal{M}) &= P(a(\epsilon) \leq \alpha, \beta \leq b(\epsilon) \,|\, \mathcal{M}) \\ &= P(\beta \leq b(\epsilon) \,|\, a(\epsilon) \leq \alpha, \mathcal{M}) P(a(\epsilon) \leq \alpha \,|\, \mathcal{M}) \\ &= \frac{1-\gamma}{1-\epsilon}(1-\epsilon) \end{aligned}$$

The last equality comes from the definition of $a(\epsilon)$ and $b(\epsilon)$. Therefore the shortest interval is obtained by taking $[a(\epsilon^*), b(\epsilon^*)]$ where

$$\epsilon^* = \text{argmin}_{\epsilon \in]0,\gamma[} \, (b(\epsilon) - a(\epsilon))$$

In practice, the values of $a(\epsilon)$ and $b(\epsilon)$ are estimated by the empirical quantiles calculated on the MCMC outputs. Using the output of the MCMC algorithm $(\tau_1^{(k)}, \ldots, \tau_n^{(k)})_{k=1,\ldots,N_{MCMC}}$, we produce $(\alpha^{(k)}, \beta^{(k)})_{t=1,\ldots,N_{MCMC}}$ the MCMC sample corresponding to the posterior distribution of (α, β),

$$\alpha^{(k)} = \min(\tau_i^{(k)},\ i \in I) \text{ and } \beta^{(k)} = \max(\tau_i^{(k)},\ i \in I).$$

$a(\epsilon)$ is estimated by the empirical quantile of the sample $\{\alpha(1), \ldots, \alpha(N_{MCMC})\}$ (sample corresponding to the marginal posterior distribution of α). $b(\epsilon)$ is estimated by the empirical quantile of $\{\beta(t), t \in T\}$ where $T = \{j \,|\, \alpha(j) \geq \epsilon\}$

2.2. Transition interval between two successive phases

The **transition interval** between two successive phases is the shortest interval that covers the end of the oldest phase and the start of the youngest phase. The start and the end are estimated by the minimum and the maximum of the dates included in the phase. It gives an idea of the transition period between two successive phases. From a computational point of view this is equivalent to the phase time range calculated between the end of the oldest phase and the start of the youngest phase.

Definition 2. *Consider a succession of two phases, P_1 and P_2. Assume P_1 is older than P_2. We denote by α_i (resp. β_i) the minimum (resp. the maximum) of the group of dates P_i $(i = 1, 2)$. The $100(1-\gamma)\%$ transition interval is the shortest interval $[a, b]$ such that*

$$P(a \leq \beta_1 \leq \alpha_2 \leq b \,|\, \mathcal{M}) = 1 - \gamma \tag{3}$$

The construction of this interval is obtained by the same step as the range time interval by replacing (α, β) by (β_1, α_2).

2.3. Testing procedure for gap between two successive phases

Successive phases may also be separated in time. Indeed there may exist **a gap** between them. This testing procedure check whether a gap exists between two successive phases with fixed probability. If a gap exists, it is an interval that covers the end of one phase and the start of the successive one with fixed posterior probability.

Definition 3. *Consider a succession of two phases, P_1 and P_2. Assume P_1 is older than P_2. We denote by α_i (resp. β_i) the minimum (resp. the maximum) of the group of dates P_i ($i = 1, 2$).*

The $100(1-\gamma)\%$ gap between these successive phases (if it exists) is the longest interval $[a, b]$ such that

$$P(\beta_1 \leq a \leq b \leq \alpha_2 \,|\, \mathcal{M}) = 1 - \gamma \tag{4}$$

The first step consists in the construction of all the couples $a(\epsilon), b(\epsilon)$ such that

$$P(\beta_1 \leq a(\epsilon)\,,\, b(\epsilon) \leq \alpha_2 \,|\, \mathcal{M}) = 1 - \gamma$$

where

- $a(\epsilon) = F^{-}_{\beta_1}(1-\epsilon \,|\, \mathcal{M})$ and $F^{-}_{\beta_1}(\cdot \,|\, \mathcal{M})$ is the quantile function of the posterior distribution of β_1
- $b(\epsilon) = F^{-}_{\alpha_2}(\frac{\gamma-\epsilon}{1-\epsilon} \,|\, \beta_1 < a(\epsilon), \mathcal{M})$ and $F^{-}_{\alpha_2}(\cdot \,|\, \beta_1 < a(\epsilon), \mathcal{M})$ is the quantile function of the conditional distribution of α_2 given $\beta_1 < a(\epsilon)$ and $\mathcal{M}$.

Indeed

$$\begin{aligned} P(\beta_1 \leq a(\epsilon)\,,\, b(\epsilon) \leq \alpha_2 \,|\, \mathcal{M}) &= P(\beta_1 \leq a(\epsilon) \,|\, \mathcal{M}) P(b(\epsilon) \leq \alpha_2 \,|\, \beta_1 \leq a(\epsilon)\,,\, \mathcal{M}) \\ &= (1-\epsilon)(1 - \frac{\gamma-\epsilon}{1-\epsilon}) = 1 - \gamma \end{aligned}$$

Testing procedure : If the set $E = \{\epsilon \,|\, a(\epsilon) < b(\epsilon)\}$ is empty, then we conclude that, with probability $1-\gamma$, no gap exists between these successive phases. Otherwise, the gap interval is the longest interval $[a(\epsilon^*), b(\epsilon^*)]$ where

$$\epsilon^* = \text{argmax}_{\epsilon \in E}\ (b(\epsilon) - a(\epsilon))$$

2.4. Tempo Plot: Rhythms of occurrence of events

Dye (2016) proposes a graphical tool, called **tempo plot**, to evaluate the rhythms of occurrence of events. We propose a Bayesian interpretation of the estimate. The quantity of interest cannot be viewed as a counting process because the date of the events $(\tau_1, \ldots, \tau_n)$ are not observed. They are estimated from a Bayesian chronological model (e.g Oxcal, ChronoModel). For each date t, the aim is to estimate the number of events $N(t)$ which occurs before the date t, we have

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

As $N(t)$ is a function of the parameters $(\tau_1, \ldots, \tau_n)$, we can easily estimate $N(t)$ from the joint posterior distribution of $(\tau_1, \ldots, \tau_n)$. The Bayes estimate of $N(t)$ (under quadratic loss) is the posterior mean of $N(t)$ i.e.

$$\begin{aligned} \hat{N}(t) &= \mathbb{E}(N(t) \,|\, \mathcal{M}) \\ &= \mathbb{E}(\sum_{i=1}^{n} \mathbb{I}_{]-\infty,t]}(\tau_i) \,|\, \mathcal{M}) \\ &= \sum_{i=1}^{n} P(\tau_i < t \,|\, \mathcal{M}). \end{aligned} \tag{5}$$

Using the output of the MCMC algorithm $(\tau_1^{(k)}, \ldots, \tau_n^{(k)})_{k=1,\ldots,N_{MCMC}}$, we can approximate the Bayes estimate $\hat{N}(t)$ by taking

$$= \sum_{i=1}^{n} \frac{1}{N_{MCMC}} \sum_{k=1}^{N_{MCMC}} \mathbb{I}_{]-\infty,t]}(\tau_i^{(k)})$$
$$= \frac{n}{N_{MCMC}} \sum_{k=1}^{N_{MCMC}} F^{(k)}(t)$$

where $F^{(k)}$ is the empirical cumulative distribution of the sample $(\tau_1^{(k)}, \ldots, \tau_n^{(k)})$ (output of the iteration number of k of MCMC algorithms)

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

The N_{MCMC} functions $nF^{(k)}(.)$ y provide a sample from the posterior distribution of N. Therefore we can easily build a credible confidence region for the function N. Indeed for each t we take the smaller posterior interval approximated from the sample $nF^{(k)}(t), k = 1, \ldots, N_{MCMC}$. An alternative is to use Gaussian approximation to get the confidence intervals.

Alternative to Tempo plot consist in representing the activity, that is the first derivative of $\hat{N}$ defined in (5).

3. The example of Ksar Akil

At Ksar 'Akil (Lebanon), a deep Paleolithic stratigraphic sequence was investigated in order to established the chronology of the site (See Bosch *et al.* (2015)). This stratigraphic sequence included Initial Upper Paleolithic (IUP), from layer XXV (the bottom of the sequence) to layer XXI, Ahmarian, from layer XX to layer XIV, Upper Paleolithic (UP), from layer XV to layer VI, and Epi-Paleolithic (EPI), from layer V to layer I. In addition, 16 shell ornaments were found throughout the stratigraphy and dated by AMS radiocarbon technique.

The aim of this modeling was to establish the chronology of the succession of phases: IUP, Ahmarian, UP and EPI using these 16 radiocarbon dates and the information coming from the stratigraphy.

For the modeling with ChronoModel, the study period was set to -50 000 to -25 000. Target events were the death of the shells. Each event was then associated with one radiocarbon date. Each layer was modeled by a phase including all events (death of shells) found within each layer in the stratigraphy. Then IUP, Ahmarian, UP and EPI phases gathered all the events of the layers included in these phases.

The Markov Chains simulated by ChronoModel were then extracted to illustrate the use of **RChronoModel**. This procedure is described in Section 4.

4. Extracting the MCMC sample of the joint distribution

4.1. From Oxcal

Once your modeling is done with Oxcal, add the MCMC_Sample command to your script.

```
MCMC_Sample([Name], [Interval], [Max]) {...} ) ;
```

This command allows all of the MCMC samples to be saved to a file (default is MCMC_Sample.csv) at defined intervals and with a maximum number of samples taken (default 1000); the file can be found as a Document

4.2. From ChronoModel

Once your modeling is done with ChronoModel and once the RUN command is launched, you may save all MCMC samples in format BC/AD.

To do that, use

on the right hand side of the window in the Results tab (See figure 1). Several CSV files will be created: a file called "events.csv" containing the MCMC samples of all events, a file called "phases.csv" containing the parameter of the minimum (alpha) and the parameter of the maximum (beta) of all phases if at least one phase is modeled, and a file per phase (if any) containing the MCMC samples of the parameters of the phase and all events included in it.

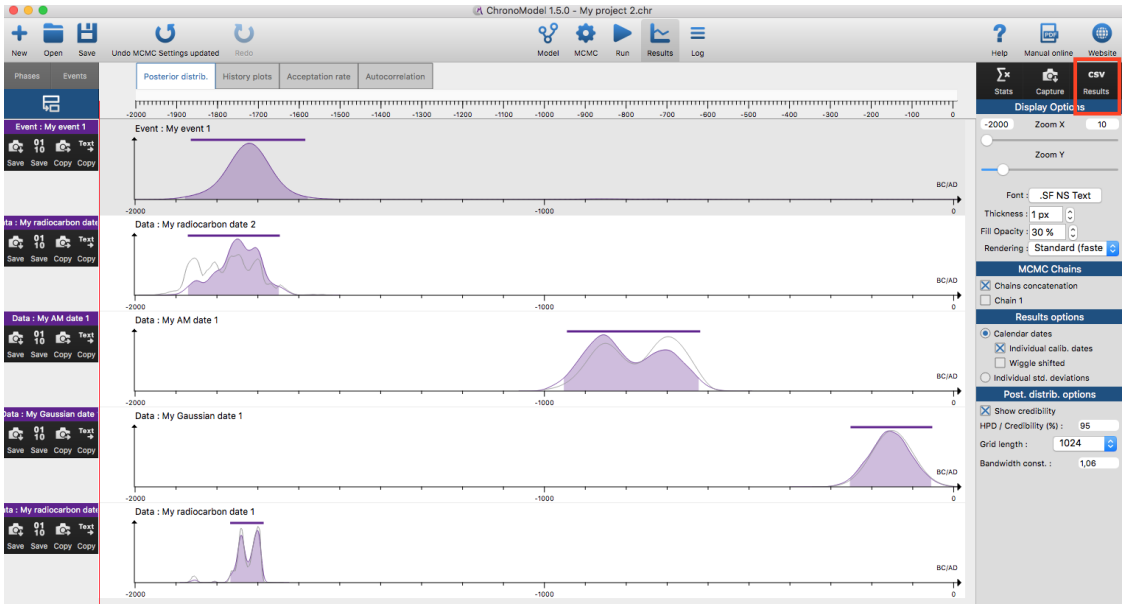

Figure 1: ChronoModel interface showing the results of a modeling.

5. The use of RChronoModel

5.1. Installing the package RChronoModel

When R is launched, you need to install (only the first time) the package **RChronoModel** using the following code:

```
R> install.packages('RChronoModel', dependencies = TRUE)
```

Then the first step is to load the library:

```
R> library(RChronoModel)
```

and then to define your working directory. Usually, the working directory is the directory where the data exported from ChronoModel or Oxcal is saved. For example, if the data are stocked in a directory called "ChronoModel/data" on the desktop of the computer, the working directory will be changed using the following command:

```
R> setwd("~/Desktop/ChronoModel/data")
```

5.2. Importing data into R software

To import the data file into R, you may use ImportCSV function. For CSV files extracted from ChronoModel sotfware, there is no need to specify any other parameters than the name of the file (and the path to it). Otherwise, you may change the specification after "sep=" and "dec=".

```
R> Phases <- ImportCSV("phases.csv", sep=",", dec=".", comment.char="#")
```

Or equivalently

```
R> Phases <- ImportCSV("phases.csv")
```

The parameter "comment.char=" is used to define how comments are written in the file to be imported. Comments of all csv files generated by ChronoModel are specified by '#'.

Now, the following code

```
R> attach(Phases)
```

makes all objects of the database "Phases" accessible by simply giving their names without having to recall the dataset of origin. For instance, the two following lines give the same result:

```
R> summary(Phases$Ahmarian.alpha)
R> summary(Ahmarian.alpha)
```

The output of the R console is the following:

```
> summary(Phases$Ahmarian.alpha)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 -45920  -41640  -41420  -41450  -41220  -40240
> summary(Ahmarian.alpha)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 -45920  -41640  -41420  -41450  -41220  -40240
```

This will be usefull in the following sections.

5.3. Examining an archeological phase

To understand chronologies, one can gather dates with common criteria into groups or phases. These groups are characterised by the date of the minimum of the group and the date of the maximum of the group. For chronologies estimated with ChronoModel, the minimum (called alpha) and the maximum (called beta) of each group of dates or each phase are computed and saved in the file called "phases.csv". However, for chronologies constructed with Oxcal, as the function MCMC_Sample() does not allow to save the MCMC samples of the minimum and the maximum (functions First() and Last()), we need to create a file containing these values. To do that, we may use the function `CreateMinMaxGroup()`.
In the following, we will use the file "phases.CSV" created and exctrated from ChronoModel.

From the output of the MCMC algorithm, one can estimate the duration of the groups as the value of the maximum - minimum at each iteration. One can also estimate the phase time range as the shortest interval that contains all the dates of the group at a given confidence level (See section 2 for the statistical details).

If interested in summary statistics of all posterior distributions related to a phase / group of dates, use the following code:

```
R> PhaseStatistics(Ahmarian.alpha, Ahmarian.beta, level = 0.95)
```

The output of the **R** console is the following:

```
> PhaseStatistics(Ahmarian.alpha, Ahmarian.beta, level = 0.95)
                        Minimum       Maximum Duration
mean                  -41447.00 -38141.00  3306.00
MAP                   -41420.00 -38149.00  3295.00
sd                       335.00    412.00   526.00
Q1                    -41639.00 -38396.00  2958.00
median                -41417.00 -38146.00  3285.00
Q2                    -41221.00 -37898.00  3631.00
level                      0.95      0.95     0.95
CredibleInterval Inf -42134.00 -38964.00  2292.00
CredibleInterval Sup -40857.00 -37317.00  4351.00
HPDRInf 1             -42080.00 -38971.00  2264.00
HPRDSup 1             -40778.00 -37320.00  4330.00
```

`PhaseStatistics()` results in a matrix of several summary statistics according to the marginal posterior density of the minimum, the maximum and the duration of the phase Ahmarian. The default confidence level is 0.95 but may be changed.

The following code gives the endpoints of the time range of the phase Ahmarian (the group of dates that constitute the phase Ahmarian) and recall the given confidence level.

```
R> PhaseTimeRange(Ahmarian.alpha, Ahmarian.beta)
```

The output of the console is

```
> PhaseTimeRange(Ahmarian.alpha, Ahmarian.beta)
       level TimeRangeInf TimeRangeSup
        0.95    -42210.62    -37319.37
```

The following lines may be used to draw a plot of the marginal posterior density of the minimum and the maximum of a phase on a same graph.

```
R> PhasePlot(Ahmarian.alpha, Ahmarian.beta, title ="Characterisation of phase Ahmarian")
```

The result is shown on Figure 2. Marginal posterior densities of the minimum and the maximum of the phase Ahmarian (curves) are presented with the shortest credible interval (solid line under the curve) at the desired level (default = 95 %) and their mean value (a dot under the curve) using the same color. In addition, the time interval of the phase at the desired level (default = 95 %) is also presented by a solid line above the curves.

The following code draws the marginal posterior density of the duration of the phase with its shortest credible interval and its mean value. By default, the confidence level is fixed at 0.95 and graphs are in color but these details may be changed. Figure 3 displays the result.

```
R> PhaseDurationPlot(Ahmarian.alpha, Ahmarian.beta, title ="Duration of Phase Ahmarian")
```

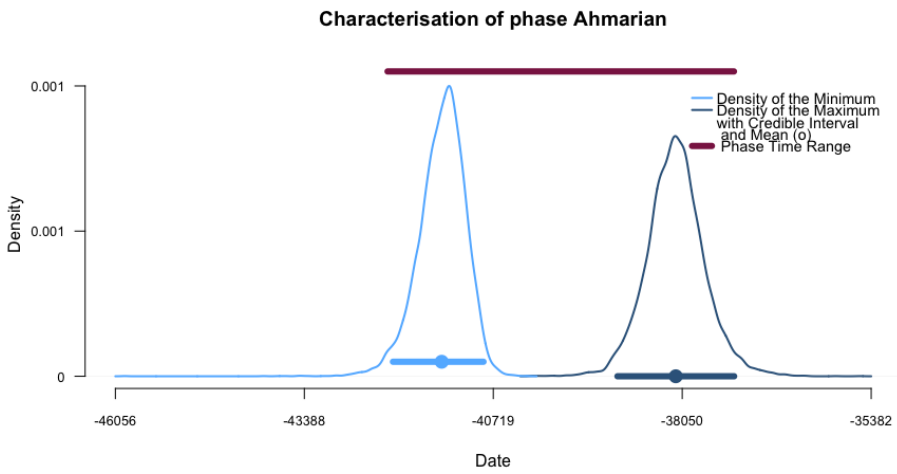

Figure 2: Plot of the marginal posterior densities of the minimum and the maximum of phase Ahmarian

5.4. Examining the succession of two phases

Let's use for this example the succession of phases "Layer XVII" and "Layer XVI", that are a part of the stratigraphic succession. As these phases are in temporal order, we can estimate, if it exists, the gap (or hiatus) and the transition interval between both phases (See section 2 for more details).

To test if there exists a gap at 95% between these successive phases, let's use the following line.

```
R> PhasesGap(Layer.XVII.beta, Layer.XVI.alpha)
```

Then the output is

```
> PhasesGap(Layer.XVII.beta, Layer.XVI.alpha)
```

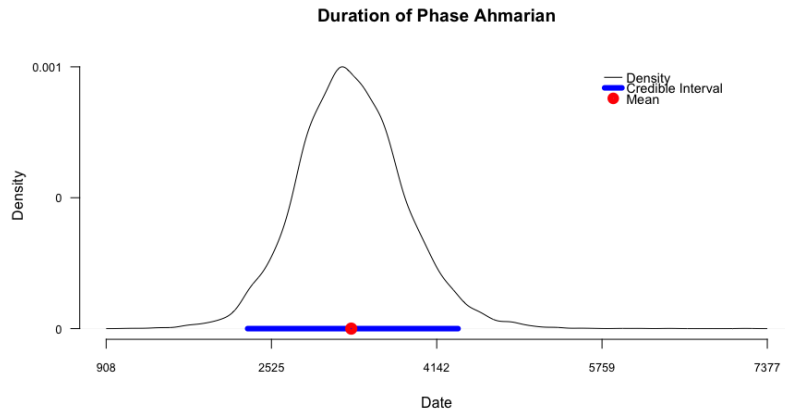

Figure 3: Plot of the marginal posterior densities related to phase Ahmarian

```
      level HiatusIntervalInf HiatusIntervalSup
     "0.95"              "NA"              "NA"
```

This code gives the endpoints of the gap interval between both phases. The default confidence level is 0.95. For this confidence level, no hiatus exists between "Layer XVII" and "Layer XVI". However, for a confidence level fixed at 0.20, there exists a gap of 50 years between both phases. To test if there exists a gap at 20% between these successive phases, let's use the following line:

```
R> PhasesGap(Layer.XVII.beta, Layer.XVI.alpha, level = 0.20)
```

Now the output is

```
> PhasesGap(Layer.XVII.beta, Layer.XVI.alpha, level = 0.2)
      level HiatusIntervalInf HiatusIntervalSup
        0.2          -38956.0          -38906.0
```

Now, to estimate the transition interval at 95% between these successive phases, let's use the following line:

```
R> PhasesTransition(Layer.XVII.beta, Layer.XVI.alpha)
```

And the result is:

```
> PhasesTransition(Layer.XVII.beta, Layer.XVI.alpha)
        0.95 TransitionRangeInf TransitionRangeSup
        0.95          -40299.00          -38043.00
```

At 95%, the transition between these two phases starts at -40 299 and ends at -38 043.
All these pieces of information may be illustrated on a graphic using the following command line :

```
R> SuccessionPlot(Layer.XVII.alpha, Layer.XVII.beta, Layer.XVI.alpha, Layer.XVI.beta)
```

Figure 4 presents the resulting graph.

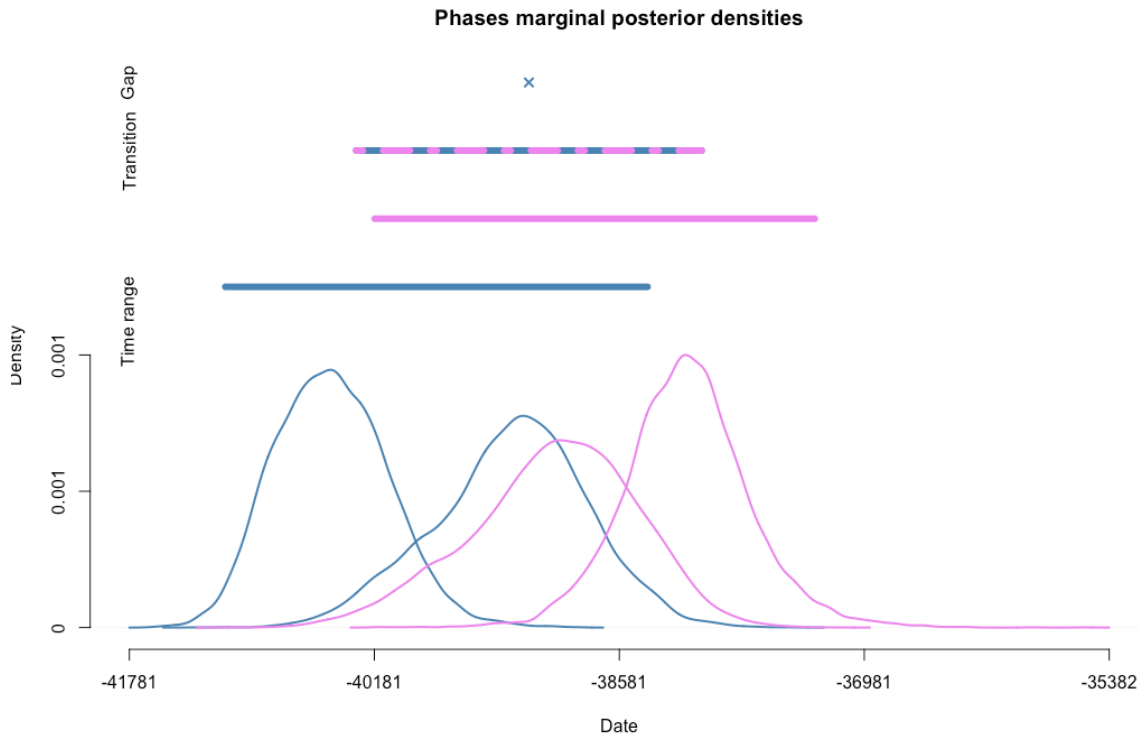

Figure 4: Plot of the succession of phases "Layer XVII" and "Layer XVI" and 95% intervals. Curves represent the densities of the minimum and the maximum of each phase. The oldest phase is drawn in blue, the youngest phase in violet. Segments correspond to time range of the phase of the same color, two-colored segments correspond to transition interval or to the gap range. A cross instead of a two-colored segment means that there is no gap range at the desired level of confidence.

5.5. Examining a series of phases

We may also be interested in a series of phases, as for instance the following phases of Ksar Akil: IUP, Ahmarian, UP and EPI. Let assume that we do not know about their temporal order succession. We could, for example, wish to know the time range of these different phases, or we could want to draw the different minimums and maximums on a same graph. We could also wish to know the credible interval of all minimums.

To do that, the "Multi" functions are available. For data extracted from ChronoModel software or using the function `CreateMinMaxGroup`, only the vector of positions of all phases' minimums are needed. Otherwise, the vector of positions of the phases' maximums is also required.

The following line gives the endpoints of the time range of phases whose minimums are in position 2, 4, 6 and 14.

```
R> MultiPhaseTimeRange(Phases, position_minimum = c(2,4,6,14))
```

By default, the argument of `position_maximimum` is equals to `position_minimum + 1`.
The result is:

```
> # Phase time ranges
```

```
> MultiPhaseTimeRange(Phases, position_minimum = c(2,4,6,14))
                              Level TimeRangeInf TimeRangeSup
EPI.alpha EPI.beta             0.95       -28983       -26951
UP.alpha UP.beta               0.95       -38501       -29284
Ahmarian.alpha Ahmarian.beta   0.95       -42211       -37319
IUP.alpha IUP.beta             0.95       -43272       -41155
>
```

Phase "EPI" whose minimum output is in column 2 has a time range = [-28 983, -26 951] at a confidence level of 95% and the phase "IUP" whose minimum output is in column 14 has a time range = [-43 272, -41 155].

The following code
`R> MultiPhasePlot(Phases, position_minimum = c(2,4,6,14), title="Phases EPI, UP, Ahmarian et IUP")` generates a plot presented in Figure 5. Curves represent the densities of the minimum and maximum of each phase. Segments above the curves correspond to the time range of the phases. Characteristics of a phase (minimum and maximum densities, time range) are drawn using the same color. When a group is defined by only one date, then the minimum equals the maximum date of the group. Hence, in that case, only one curve is presented.

`R> MultiCredibleInterval(Phases, position_minimum = c(2,4,6,14))`
computes the credible interval for the minimum of each of these phases. The result is:

```
> # Credible intervals
> MultiCredibleInterval(Phases, c(2,4,6,14))
               Level CredibleIntervalInf CredibleIntervalSup
EPI.alpha       0.95            -28974.5            -26949.9
UP.alpha        0.95            -38519.7            -36685.7
Ahmarian.alpha  0.95            -42134.5            -40856.9
IUP.alpha       0.95            -43268.8            -41154.1
>
```

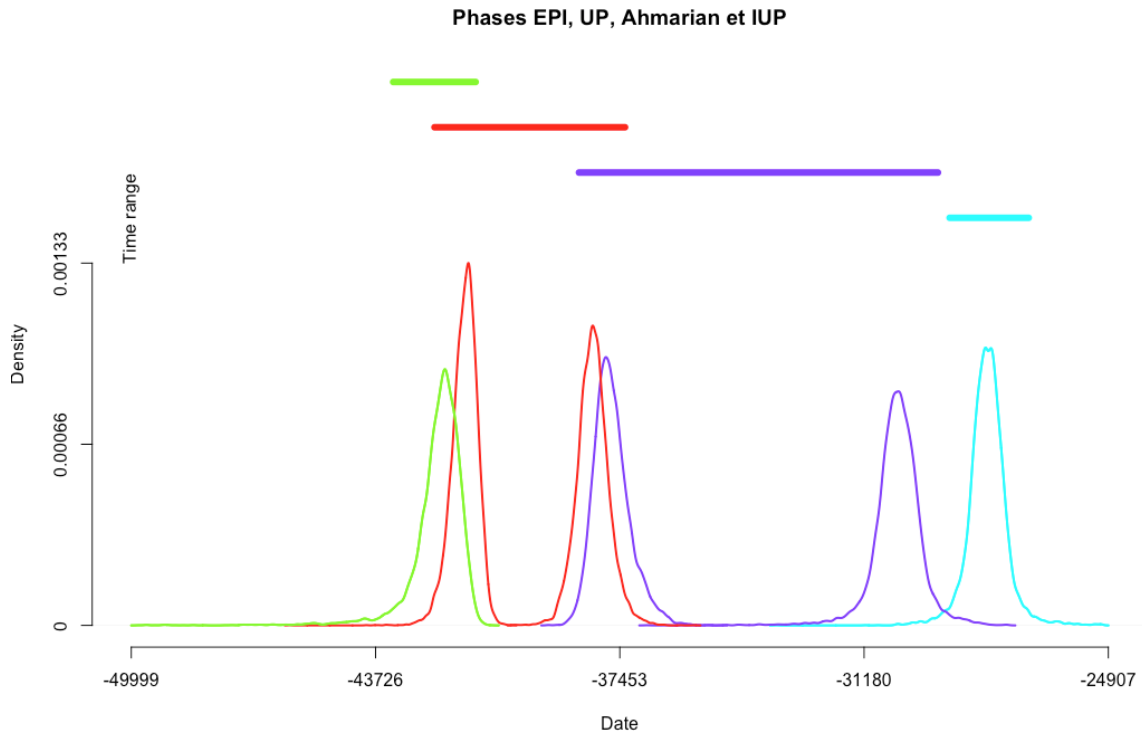

Figure 5: Plot of the series of Ksar Akil's phases and their associated time range at 95%. The marginal posterior densities of phase IUP are drawn in green, those of phase Ahmarian are in red, those of phase UP are in purple and those of phase EPI are in light blue. As there is only one date in the phases EPI and IUP, the minimum and the maximum of these phases have the same value at each iteration. Hence, we can see only one curve for each of these phases. Time ranges are displayed by segments above the curves of the same color as the densities of the phase.

5.6. Examining a succession of phases in temporal order

We may also be interested in a succession of phases. This is actually the case of the succession of IUP, Ahmarian, UP and EPI that are in stratigraphic order. Hence, we can estimate the transition interval and, if it exists, the gap between these successive phases. This may be done using the following code:

```
R> MultiPhasesGap(Phases, position_minimum=c(14,6,4,2))
R> MultiPhasesTransition(Phases, position_minimum=c(14,6,4,2))
```

For these functions, the order of the phases is important. The vector of positions of the minimums should start with the minimum of the oldest phase and end with the one of the youngest phase. For data extracted from ChronoModel, the vector of positions of the phases' maximums is deduced from the vector of the minimum. For other data, this vector should be specified.

```
> MultiPhasesGap(Phases, position_minimum = c(14,6,4,2))
                          Level  HiatusIntervalInf HiatusIntervalSup
IUP.beta & Ahmarian.alpha "0.95" "NA"              "NA"
Ahmarian.beta & UP.alpha  "0.95" "NA"              "NA"
UP.beta & EPI.alpha       "0.95" "-29202"          "-29110"
>
> # Phases Transition
> MultiPhasesTransition(Phases, position_minimum = c(14,6,4,2))
                          0.95 TransitionRangeInf TransitionRangeSup
IUP.beta & Ahmarian.alpha 0.95             -43272             -40704
Ahmarian.beta & UP.alpha  0.95             -39042             -36686
UP.beta & EPI.alpha       0.95             -31553             -26965
>
```

At a confidence level of 95%, there is no gap between the succession of phases IUP, Ahmarian and UP, but there exists one of 88 years between phase UP and phase EPI.

Figure 6 presents the resulting plot of the following lines:

```
R> MultiSuccessionPlot(Phases, position_minimum =c(14,6,4,2), title = "Ksar Akil
- Succession of phases:  IUP, Ahmarian, UP, EPI ")
```

In that plot, curves represent the densities of the minimum and maximum of each phase. When a group is defined by one only date, the minimum equals the maximum date of the group. Hence, in that case, only one curve is presented. Segments above the curves correspond to time range of the phase associated to a level confidence of 95%. Two-coloured segments correspond to transition interval or to the gap range associated to a level confidence of 95%. A cross instead of a two-coloured segment means that there is no gap between the succession of phases at the desired level of confidence.

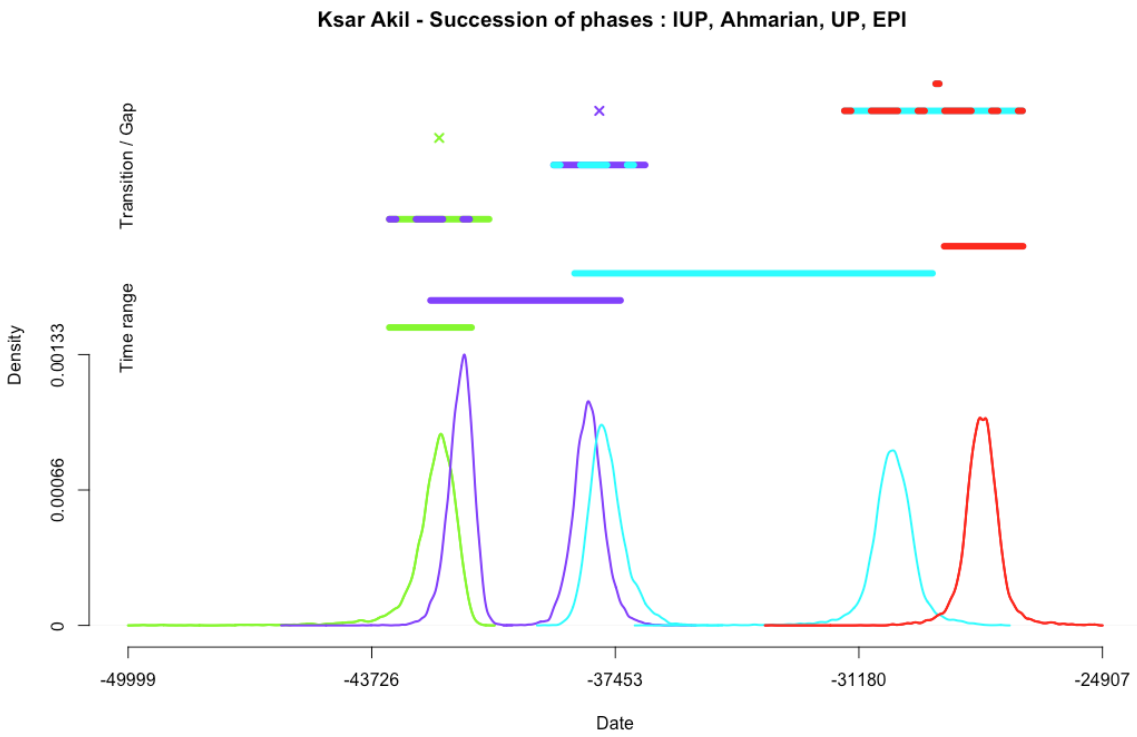

Figure 6: Plot of the succession of phases from Ksar Akil. The characteristics of phase IUP are drawn in green, those of phase Ahmarian are in purple, those of phase UP are in light blue and those of phase EPI are in red. Again, as there is only one event in the phases EPI and IUP, the minimum and the maximum of these phases have the same values at each iteration. Hence, we can only see one curve for each of these phases. Time range are displayed by segments above the curves. Two-coloured segments correspond to transition interval or to the gap range associated to a level confidence of 95%. As there are no gaps at 95% between phases IUP and Ahmarian, and Ahmarian and UP, a cross is drawn instead.

5.7. Summary

We can summary the characteristics of the different phases of Ksar Akil, using the data published by Bosch et al. Bosch *et al.* (2015) and the modeling done with ChronoModel, by the following table :

Phase	**Time range**		**Transition range**		**Gap range (if it exists)**	
	Inf	**Sup**	**Inf**	**Sup**	**Inf**	**Sup**
IUP	-43 272	-41 155	NA	NA	NA	NA
Ahmarian	-42 211	-37 319	-43 272	-40704	NA	NA
UP	-38 501	-29 284	-39 042	-36 686	NA	NA
EPI	-28 983	-26 951	-31 553	-26 965	-29 202	-29 110

Table 1: Results associated with a level of confidence at 95%

5.8. Examining a series of dates

Let's now analyse the series of dates included in this example. First, let's import the CSV file called "events.csv" that contains the resulting MCMC values of each date.

```
R> Events = ImportCSV("events.csv")
```

This file contains the MCMC samples of each of the 16 dates of the death of the 16 shells found at Ksar Akil. The name given to thoses dates are the name of the layer in which they where found. Is several where found in a same layer, then a number was associated to the name of the Layer.

Using the following code, we can estimate the 95% credible interval of each date.

```
R> MultiCredibleInterval(Events, c(2:17))
```

That gives the following results:

```
> MultiCredibleInterval(Events, c(2:17))
             Level CredibleIntervalInf CredibleIntervalSup
Layer.V       0.95              -29063              -27022
Layer.XVI     0.95              -39251              -37570
Layer.VI      0.95              -31512              -29102
Layer.XVI.1   0.95              -40013              -38378
Layer.XI      0.95              -37358              -35037
Layer.XVI.2   0.95              -40335              -37845
Layer.XVI.3   0.95              -40323              -37830
Layer.XII     0.95              -38574              -36798
Layer.XVII    0.95              -40986              -39044
Layer.XVIII   0.95              -41404              -40118
Layer.XVII.1  0.95              -40860              -39006
Layer.XIX     0.95              -41608              -40498
Layer.XVII.2  0.95              -40742              -39601
Layer.XX      0.95              -42148              -40868
Layer.XVII.3  0.95              -41198              -39616
Layer.XXII    0.95              -43307              -41114
```

The following code gives an estimation of the 95% HPD regions of each date.

```
R> MultiHPD(Events, c(2:17))
```

All those intervals may be drawn on a graph using the following line:

```
R> MultiDatesPlot(Events, c(2:17), intervals="CI", title=" 95% CI of Ksar Akil
dates")
```

Figure 7 shows the resulting graph.

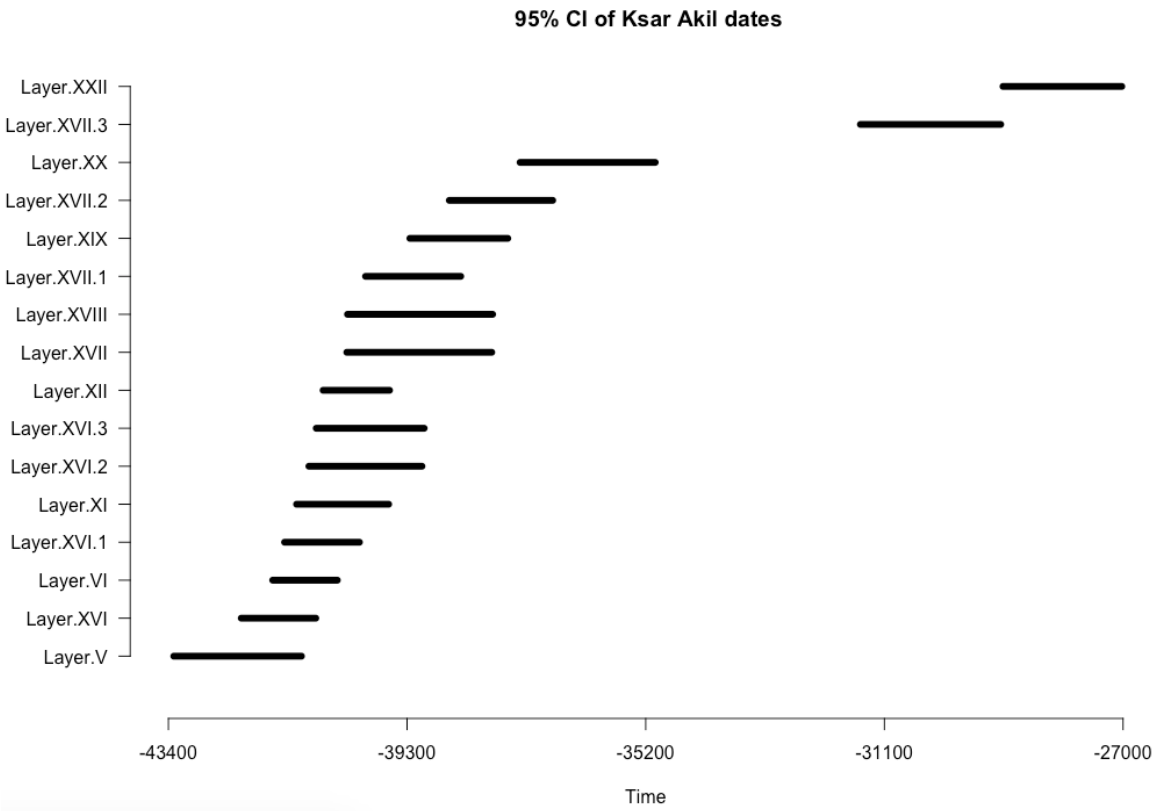

Figure 7: Graph of the credible interval of the series of dates included in Ksar Akil

Similarly, the following line gives the graph of a series of HPD regions presented in Figure 8.

```
R> MultiDatesPlot(Events, c(2:17), intervals="HPD", title=" 95% HPD regions of
Ksar Akil dates")
```

Now, the rhythm of occurrence of the dates may be investigated using two plots

```
R> TempoPlot(Events, c(2:17), title=" Tempo plot")
R> TempoActivityPlot(Events, c(2:17))
```

Figure 9 displays the tempo plot and the activity plot of the site of Ksar Akil.

From these graphs, we can see that the highest part of the sampled activity is dated between -45 000 to -35 000 but two dates are younger, at about -32 000 and -28 000.

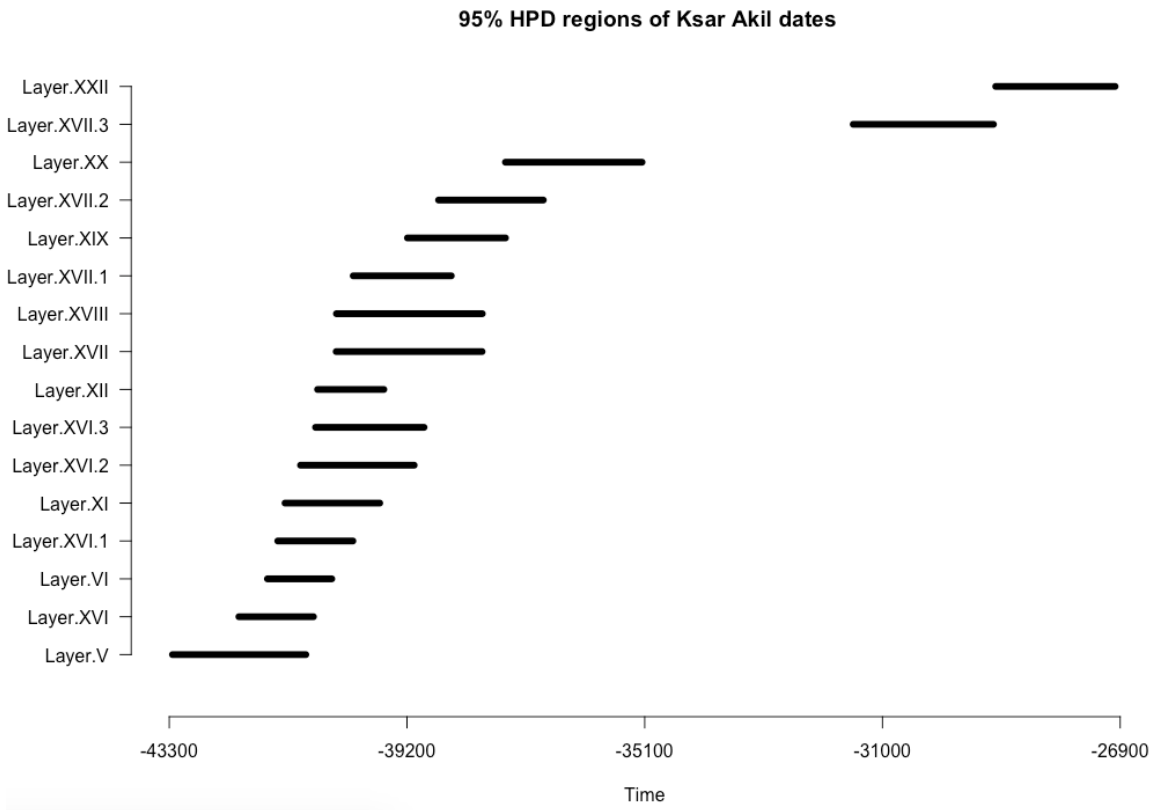

Figure 8: Graph of the HPD regions of the series of dates included in Ksar Akil

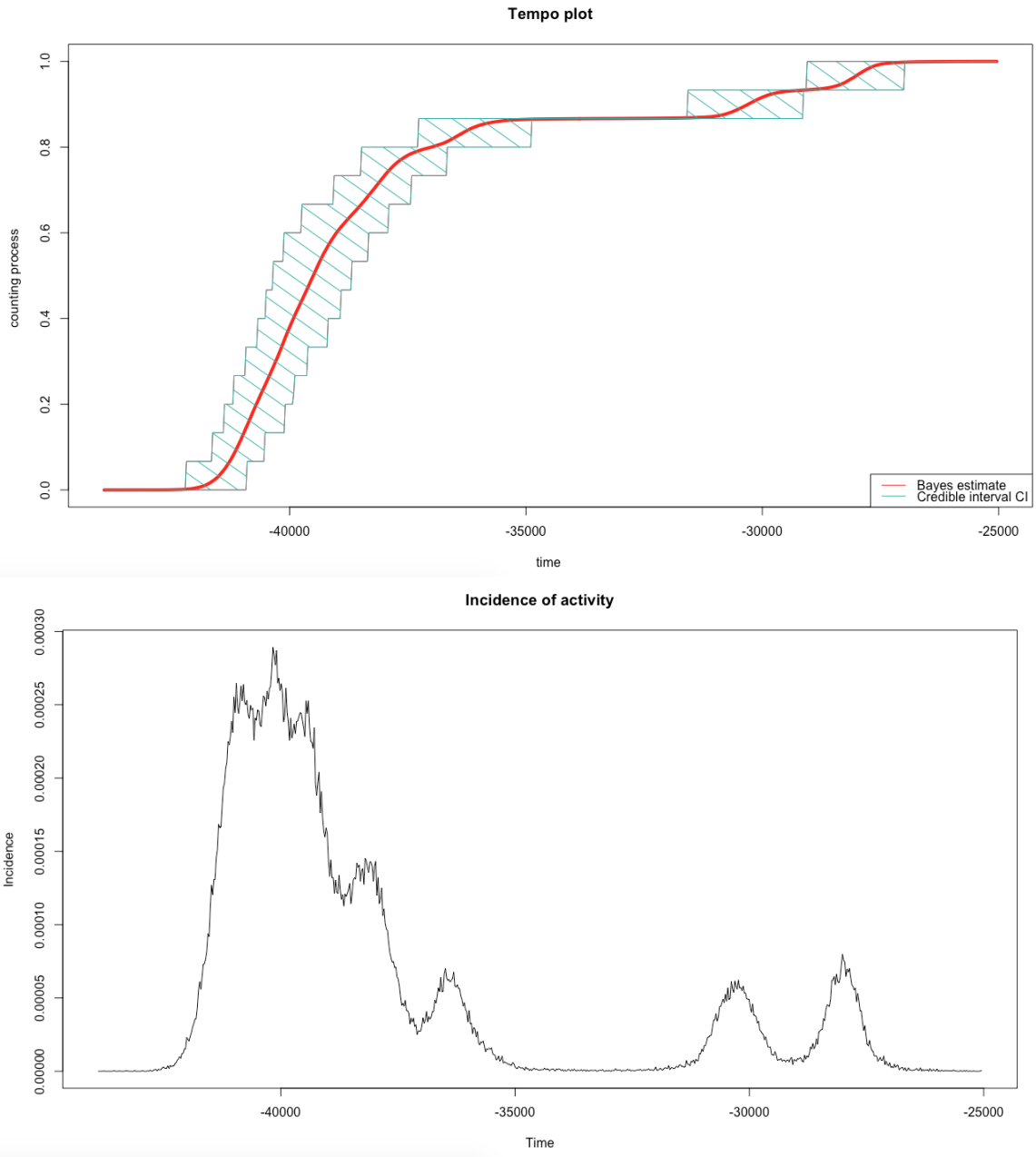

Figure 9: Tempo plot [top] and activity plot [bottom] of Ksar Akil

References

Bosch MD, Mannino MA, Prendergast AL, O'Connell TC, Demarchi B, Taylor SM, Niven L, van der Plicht J, Hublin JJ (2015). "New chronology for Ksâr âĂŸAkil (Lebanon) supports Levantine route of modern human dispersal into Europe." *Proceedings of the National Academy of Sciences*, **112**(25), 7683–7688. ISSN 0027-8424. doi:10.1073/pnas.1501529112.

Dye TS (2016). "Long-term rhythms in the development of Hawaiian social stratification." *Journal of Archaeological Science*, **71**, 1–9. doi:10.1016/j.jas.2016.05.006.

Lanos P, Philippe A, Lanos H, Dufresne P (2016). *Chronomodel : Chronological Modelling of Archaeological Data using Bayesian Statistics (Version 1.5).* URL http://www.chronomodel.fr.

Ramsey CB (2009). "Bayesian Analysis of Radiocarbon Dates." *Radiocarbon*, **51**(1), 337–360. ISSN 0033-8222.

Ramsey CB (2016). *Oxcal 4.2.* URL http://c14.arch.ox.ac.uk/oxcalhelp/hlp_contents.html.

Vibet MA, Philippe A, Lanos P, Dufresne P (2016). *ChronoModel V1.5 User's manual.* URL http://www.chronomodel.fr.

Affiliation:

Anne Philippe
Université de Nantes, Laboratoire de mathé matiques Jean Leray
2, Rue de la Houssiniere, 44000 Nantes, FRANCE
E-mail: anne.philippe@univ-nantes.fr
URL: http://www.math.sciences.univ-nantes.fr/~philippe/info.html

Marie-Anne Vibet
Université de Nantes, Laboratoire de mathé matiques Jean Leray
2, Rue de la Houssiniere, 44000 Nantes, FRANCE
E-mail: marie-anne.vibet@univ-nantes.fr