



**HAL**  
open science

# Analysis of Archaeological Phases using the CRAN Package ArchaeoPhases

Anne Philippe, Marie-Anne Vibet

► **To cite this version:**

Anne Philippe, Marie-Anne Vibet. Analysis of Archaeological Phases using the CRAN Package ArchaeoPhases. 2018. hal-01347895v5

**HAL Id: hal-01347895**

**<https://hal.science/hal-01347895v5>**

Preprint submitted on 20 Sep 2018

**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 **ArchaeoPhases**

Anne Philippe  
Université de Nantes, LMJL

Marie-Anne Vibet  
Université de Nantes, LMJL

---

## Abstract

We propose new statistical tools to analyse and to estimate the characteristics of time periods based on the posterior distribution of the associated collection of dates. These tools are implemented in the R package **ArchaeoPhases**. The required inputs are simulated samples from the posterior distribution of a collection of dates. Such Markov Chains Monte Carlo samples are provided, for instance, by BCal, ChronoModel or OxCal, freely available software applications build for the chronological modeling of estimated dates. We give a practical introduction to the package **ArchaeoPhases** using published data and comment the statistical results.

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

---

## 1. Introduction

Statistical modelling within the Bayesian framework is widely used for constructing chronologies based on relative and absolute chronological information. This consists in the estimation of a collection of dates where

- the observations are measurements (along with their error) coming from possibly different dating methods (radiocarbon, archaeomagnetism, thermoluminescence...),
- the prior information about these dates is induced by archaeological, 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 *the posterior distribution*. Usually the analytic expression of the posterior distribution is not easily computable. However, inferences may be drawn using a sample generated from the posterior distribution. 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.

Software applications available to estimate chronologies include: BCal (see [Buck, Boden, Christen, James, and Sonnenwald \(2018\)](#); [Buck, Christen, and James \(1999\)](#)), Chronomodel (see [Lanos, Philippe, Lanos, and Dufresne \(2016\)](#); [Vibet, Philippe, Lanos, and Dufresne \(2016\)](#); [Lanos and Philippe \(2017, 2018\)](#)) and OxCal (see [Bronk Ramsey \(2016, 2009\)](#)), but

also JAGS via the package **rjags** or via the package **ArchaeoChron** (see Philippe and Vibet (2017)), and STAN via the package **rstan**. These applications 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 of MCMC simulations from the joint posterior distribution of the dates. These simulated samples can be imported into R. Hence, we can explore the joint distribution of a collection of dates using the **ArchaeoPhases** package.

The **ArchaeoPhases** package provides new statistical tools for the analysis of chronologies based on the joint distribution of estimated dates. Two first tools explore, for a collection of dates, the rate of occurrence of these dates (`TempoPlot()` et `TempoActivityPlot()`). Three other tools summarise periods of time, or what we call *phases*, defined by some archaeological, paleo-environmental, geographical... criterion. A *phase* is associated with a collection of calendar dates estimated from the chronological model.

In this article, we will explain and illustrate the use of the following new statistical tools:

- **Tempo plot** introduced by Thomas Dye (see Dye (2016)), it 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 **its associated Activity plot**.
- **Time range interval** : a time interval that characterizes the period of time during which a *phase* happened.
- **Transition range interval** : a time interval that characterizes the period of time during which a transition between two successive *phases* happened.
- **Testing procedure to check the presence of a gap** between two successive *phases*. A **gap range interval** is estimated, if we accept its existence, and characterizes the hiatus between two successive *phases*.

The CRAN package **ArchaeoPhases** presents a list of functions for calculating these statistics. The inputs of these R functions are samples simulated from the posterior distribution of dates (e.g. the Markov Chains simulated by BCal, Chronomodel or OxCal). For non-R users, a web application has 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> and within the ArchaeoPhases package (`app_ArchaeoPhases()`).

In Section 2, we describe the statistical aspects of our new tools developed for the analysis of *phases* (or their associated collection of dates) in a Bayesian framework. Section 3 provides a short description of the data published by Bosch, Mannino, Prendergast, O’Connell, Demarchi, Taylor, Niven, van der Plicht, and Hublin (2015) used to illustrate the use of **ArchaeoPhases** and explanations about the problematic. Then we construct a Bayesian model with Chronomodel in order to estimate four Palaeolithic periods based the information of the stratigraphic sequence and on radiocarbon dates found at Ksâr ’Akil (Lebanon). In Section 4, we give a practical introduction to the first steps with **ArchaeoPhases** (version 1.3). Then in Section 5, we analyse the MCMC output of the modelling done with Chronomodel and estimate the four Palaeolithic periods thanks to the information found at Ksâr ’Akil.

## 2. Statistical aspects

We define a chronology as a collection of calendar dates  $\tau_1, \dots, \tau_n$ . Usually, the dates  $\tau_1, \dots, \tau_n$  are not observed, they are estimated from a Bayesian chronological model that includes relative and absolute chronological information (e.g BCal, ChronoModel, OxCal). The relative information relates to the temporal ordering of dates whereas the absolute information usually arises from possibly different dating methods.

**Notation:** We denote by  $\mathcal{M}$  the set of measurements coming from dating methods. We assume that MCMC samples from the joint posterior distribution  $p(\tau_1, \dots, \tau_n | \mathcal{M})$  of all the dates  $\tau_1, \dots, \tau_n$  are available.

Two types of questions may arise when considering a chronology of dates : the rate of occurrence of these dates and the periods of time during which these collections of dates happened. We propose new statistical tools build as post-processing steps of the MCMC sample of the chronology in order to give answers to these questions.

### 2.1. Rate of occurrence of dates

Dye (2016) defined a graphical tool, called **tempo plot**, to evaluate the rate of occurrence of the dates. We propose a Bayesian interpretation of this estimate. The quantity of interest cannot be viewed as a counting process because the date of the events ( $\tau_1, \dots, \tau_n$ ) are not observed. For each date  $t$ , the aim is to estimate the number of dates  $N(t)$  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  $(\tau_1, \dots, \tau_n)$ , we can easily estimate  $N(t)$  from the joint posterior distribution of  $(\tau_1, \dots, \tau_n)$ . The Bayes estimate of  $N(t)$  (under quadratic loss) is the posterior mean of  $N(t)$  i.e.,

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

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

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

where  $F^{(k)}$  is the empirical cumulative distribution of the sample  $(\tau_1^{(k)}, \dots, \tau_n^{(k)})$  (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  $nF^{(k)}(\cdot)$  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  $nF^{(k)}(t), k = 1, \dots, N_{MCMC}$ . An alternative is to use Gaussian approximation to get the confidence intervals. To visualize the rhythms, an alternative to the Tempo Plot consists in representing the first derivative of  $\hat{N}$  defined in (1). This graphical tool is called Activity Plot.

## 2.2. Periods of time

The question about a chronology is not only a matter of estimation of these calendar dates  $\tau_1, \dots, \tau_n$  but also the characterization of historical, or geological, or archaeological... periods of time, or what we call *phases*. A *phase* is a collection of dates estimated from the chronology. We define a *phase* (denoted  $P$ ) by a collection of dates gathered on some criterion,  $\tau_i$  for  $i \in I$  where  $I$  is a subset of  $\{1, \dots, n\}$  and  $n$  the total number of calendar dates.

In this section, we present the classical way of summarising a periods of time, or a *phase*, and we introduce our new concepts whose aim are to further interpret the MCMC sample from the joint distribution of the chronology of dates.

### *Classical summary*

A *phase* is commonly summarized by two parameters: the start and the end dates of the corresponding period of time. Two different approaches are considered in order to estimate these two dates.

1. The start and end dates are modelled, and so additional parameters are incorporated in the model. This approach is implemented in OxCal application where the parameters are denoted  $t_a$  and  $t_b$ . This requires to do prior assumption on the distribution of the dates belonging to the *phase* on the estimated period  $[t_a, t_b]$ , such as a uniform distribution on  $[t_a, t_b]$ . See Naylor and Smith (1988); Buck, Litton, and Smith (1992); Christen (1994); Buck, Litton, and Cavanagh (1996); Nicholls and Jones (2002) for different choice of prior distributions.
2. The start (resp. the end) date is estimated by the date of the minimum (resp. the maximum) of the associated collection of dates. This approach does not model any additional parameter to define the *phase*. It is only a reflection of the time period covering the collection of dates, a post-processing step based on the MCMC sample of the collection of dates. These statistics are implemented in ChronoModel and OxCal applications.

With both approaches, a *phase* is characterised by two dates (start and end) for which a credible interval may be given. However, these two estimated parameters do not provide a localisation of the *phase* in time. For instance, one could use the Bayes estimate of the start

and end dates of a *phase* in order to build an interval. However, no one knows what is the probability that all the dates of the collection belong to that interval.

Our new tools are build in order to characterise a *phase* by a time interval with fixed probability that the dates of the collection belong to it.

### *Time Range interval*

We propose to characterize a *phase* by a time period that contains all the dates of the associated collection with a given probability. The so-called **time range interval** gives an idea of the start, the end and the duration of the *phase*. It locates a *phase* in time.

**Definition 1.** Let  $P = \{\tau_i \mid i \in I \subset \{1, \dots, n\}\}$  be a collection of dates gathered in a *phase*. The  $100(1 - \gamma)\%$  time range is the shortest interval  $(a, b)$  such that

$$P(\tau_i \in [a, b], \forall i \in I \mid \mathcal{M}) = 1 - \gamma \quad (2)$$

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

We denote  $\alpha = \text{start}(\tau_i, \forall i \in I)$  and  $\beta = \text{end}(\tau_i, \forall i \in I)$ . Equation (2) can also be rewritten in the form

$$P(a \leq \alpha, \beta \leq b \mid \mathcal{M}) = 1 - \gamma \quad (3)$$

Therefore the construction of the time range interval depends only on the joint posterior distribution of  $(\alpha, \beta)$ .

The intervals  $[a, b]$  satisfying Equation (3) are those that satisfy:

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

where

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

So we have

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

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

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

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  $(\tau_1^{(k)}, \dots, \tau_n^{(k)})_{k=1, \dots, N_{MCMC}}$ ,

we produce  $(\alpha^{(k)}, \beta^{(k)})_{t=1, \dots, N_{MCMC}}$  the MCMC sample corresponding to the posterior distribution of  $(\alpha, \beta)$ ,

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

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

#### *Transition interval between two successive phases*

We propose to characterize the transition of time between two successive *phases*, for *phases* in temporal order relationship, by a **transition interval**. This tool gives an estimation the period of time during which the older *phase* has not yet ended and the younger *phase* has just begun.

For *phases* in temporal order relationship, the **transition range interval** between two successive *phases* is the shortest interval that covers the end date of the older *phase* and the start date of the younger *phase*. From a computational point of view this is equivalent to the *time range interval* calculated between the end of the older *phase* and the start of the younger *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  $\alpha_i$  (respectively  $\beta_i$ ) the start date (respectively the end date) of the collection of dates  $P_i$  ( $i = 1, 2$ ). The  $100(1 - \gamma)\%$  transition interval is the shortest interval  $[a, b]$  such that

$$P(a \leq \beta_1 \leq \alpha_2 \leq b \mid \mathcal{M}) = 1 - \gamma \quad (4)$$

The construction of this interval is obtained the same way as the *time range interval* by replacing  $(\alpha, \beta)$  by  $(\beta_1, \alpha_2)$ .

#### *Testing procedure for gap between two successive phases*

*Phases* in temporal order relationship may be separated in time by a **gap**, or a **hiatus**. We propose a testing procedure in order to check whether a gap of time exists between two successive *phases* with fixed probability.

For *phases* in temporal order relationship, if a gap exists, it is the longest interval in between the end date of the older *phase* and the start date of the youngest 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  $\alpha_i$  (respectively  $\beta_i$ ) the start date (respectively the end date) of the collection 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

$$P(\beta_1 \leq a \leq b \leq \alpha_2 \mid \mathcal{M}) = 1 - \gamma \quad (5)$$

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

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

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

Indeed

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

**Testing procedure** : If the set  $E = \{\epsilon | a(\epsilon) < b(\epsilon)\}$  is empty, then we conclude that no gap exists between these successive *phases*, with probability  $1 - \gamma$ . Note that the probability  $1 - \gamma$  cannot be interpreted as a confidence level. But it means that we cannot construct an interval satisfying (5) with posterior probability  $1 - \gamma$ . Otherwise, the gap interval is the longest interval  $[a(\epsilon^*), b(\epsilon^*)]$  where

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

### 3. Bayesian modeling for the stratigraphy of Ksâr 'Akil

We illustrate the use of the **ArchaeoPhases** package and mainly the functions described in Section 2 with the construction and the estimation of a Palaeolithic chronology based on the stratigraphic information and measurements made on the site of Ksâr 'Akil (Lebanon).

We estimate the Palaeolithic chronology thanks to a Bayesian model implemented in **ChronoModel** (see Vibet *et al.* (2016)). This software application has a user-friendly graphical interface easily manipulated in order to define the Bayesian hierarchical model developed in Lanos and Philippe (2017, 2018). As this application is not connected with R, we should first construct the model with **ChronoModel**, then extract from this application the MCMC sample of the joint posterior distribution, and finally import the MCMC sample into R application and apply the **ArchaeoPhases** functions.

#### 3.1. The stratigraphy of Ksâr 'Akil

At Ksâr 'Akil (Lebanon), a deep Palaeolithic stratigraphic sequence was investigated in order to establish the chronology of the site (see Bosch *et al.* (2015)). This stratigraphic sequence is constituted of a succession of thirty-six layers. Based on the lithic assemblages found throughout the sequence, twenty-five layers were associated with four Palaeolithic deposits : Initial Upper Palaeolithic (IUP), from layer XXV (the bottom of the sequence) to layer XXI; Ahmarian, from layer XX to layer XIV; Upper Palaeolithic (UP), from layer XV to layer VI; and Epi-Palaeolithic (EPI), from layer V to layer I. The aim of this modelling is to establish the chronology of these four successive Palaeolithic periods (IUP, Ahmarian, UP and EPI) thanks to the chronology of the stratigraphy.

Sixteen consumed shellfish were found throughout the stratigraphy and dated by AMS radiocarbon technique in order to estimate the date of their death. Dating the death of these consumed shellfish, whose shells were directly discarded after consumption, gives an indication



of human activity on the site. Hence it gives also an indication of the different Palaeolithic periods during associated with this human activity. Thus, the chronology of the site was established on these evidences of human activity and on the stratigraphic information.

### 3.2. Modelling with Chronomodel

Based on the information given by Bosch *et al.* (2015), we propose a Bayesian model for the stratigraphy of Ksâr 'Akil implemented in ChronoModel (version 1.5).

In this model, the study period is set from -50 000 to -10 000 (dates are in calendar years Before Common Era) out of historical information on the Upper Palaeolithic.


As the aim of this modelling is to date the human occupations of the site, sixteen molluscan species collected for consumption and whose shells were directly discarded after consumption were dated by radiocarbon. This technique estimates the date of the death of these mollusks. Hence, dated events (death of mollusk) and target events (evidence of human activity) are nearly identical (see Dean (1978)). We use these dated events in order to date human activity in the layer in which these food remains were found.

Figure 1 presents the modelling realised with ChronoModel application. Dated events (death of mollusk) are called according to the name given by the laboratory (for instance : GrA-53081). Each dated event is a proxy of the human occupation in the layer in which the mollusk has been found. Target events (evidence of human activity) are named according to the associated layer. In the best cases, several mollusks were found in the same layer, and so all of these dates give an estimation of the period of time corresponding to human activity in that layer. Stratigraphic constraints, that give a temporal order to the occupation of the different layers, are also included in the model, they are symbolised by black arrows in Figure 1.

Then, layers and the 16 related events, are gathered in Palaeolithic periods in order to give an estimation of these Palaeolithic periods. In ChronoModel, four *phases* were constructed based on the evidences of human activity found in several layers. Figure 1 shows the stratigraphy of the layers in which sellfish were found, and the four Palaeolithic deposits.

### 3.3. Extracting the MCMC sample of the joint distribution

For this Bayesian modelling, 3 Markov chains are run in parallel. For each chain, 1 000 iterations are generated during the Burn-in period, 20 batches of 500 iterations are used in the Adapt period, 100 000 iterations are drawn in the Acquire period by only 1 out of 10 are kept in order to break the correlation structure.

Now, the Markov chains simulated by ChronoModel have to be extracted and then imported into R in order to use the **ArchaeoPhases** package. To do that, use  on the right hand side of the window in the Results tab (see Figure 2).

Several CSV files are created: a file called "events.csv" containing the MCMC samples of all events (target events), a file called "phases.csv" containing the date of the minimum (called alpha) and the date of the maximum (called beta) of each phase, if at least one *phase* is modelled, and a file per *phase* (if any) containing the MCMC samples of the parameters of the *phase* (the minimum and the maximum date) and all events included in it. Note that the minimum (resp. maximum) of a group of dates is an estimation of the start (resp. end) of the group of dates.

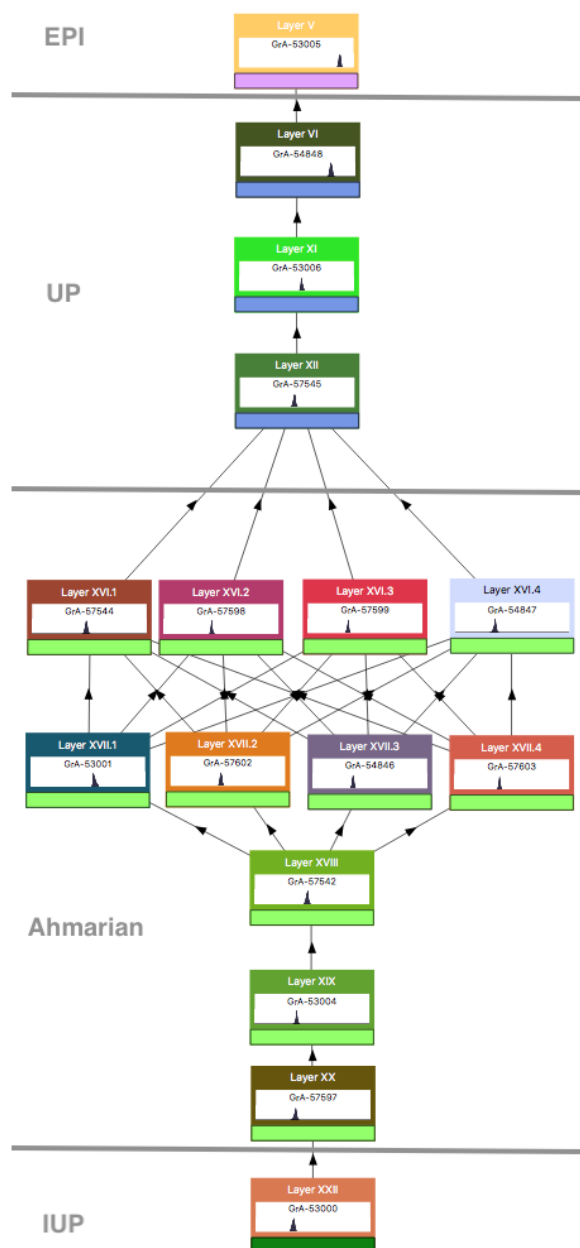


Figure 1: ChronoModel interface showing the modeling step of the chronology. Rectangular nodes represent the evidences of human activity (target events) associated with the radiocarbon date of the death of a mollusk (dated events). An arrow between two rectangular nodes symbolises a time order relationship. Gray lines symbolise the four Palaeolithic deposits and the layers associated with these deposits.

All those files may be imported into R in order to use the **ArchaeoPhases** package.

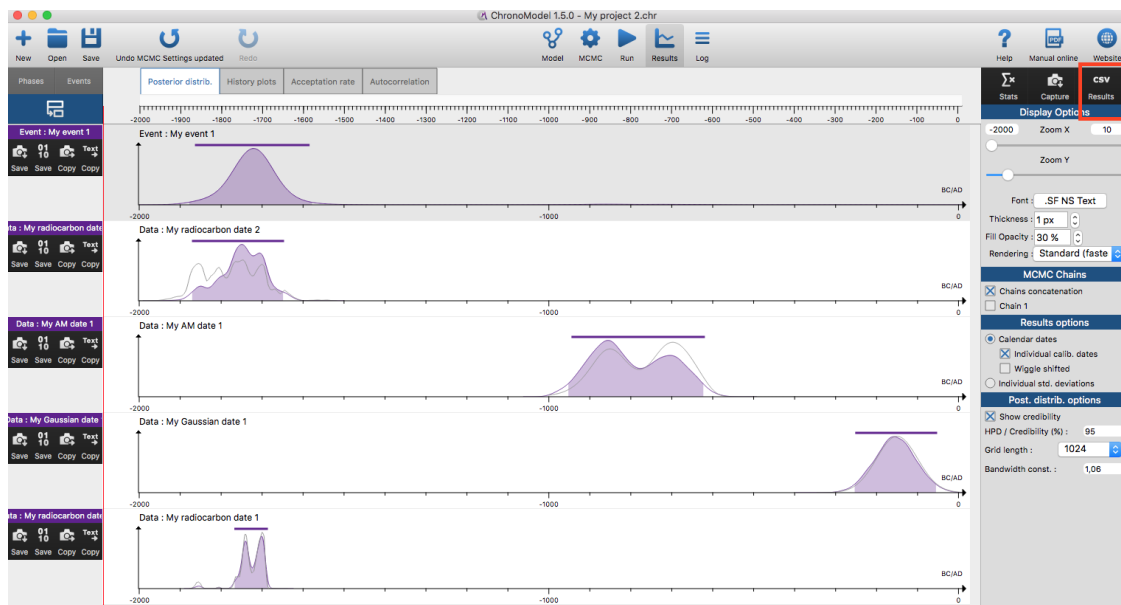


Figure 2: ChronoModel interface showing the posterior distribution of a target event with associated dated events. The icon CSV (red box) should be used to export cvd file containing MCMC samples

## 4. First steps with ArchaeoPhases

### 4.1. Installing the ArchaeoPhases package

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

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

Then the first step is to load the library:

```
R> library("ArchaeoPhases")
```

### 4.2. Importing data into R software

To import a data file into R, you may use `ImportCSV()` function. For CSV files extracted from ChronoModel software, 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=`. 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 '#'. For example,

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

In the package **ArchaeoPhases**, a dataset containing the MCMC samples of the posterior distribution of the 16 events identified in Ksâr 'Akil , and generated by ChronoModel application as explained in the previous section, is already included.

```
R> data(KADatesChronoModel)
```

This dataset contains 3 Markov chains run in parallel. The first column corresponds to the iteration number, the following ones correspond to the posterior distribution of the 16 target events.

### 4.3. Diagnostic tools

Before any analysis, the convergence of the MCMC sample should always be checked (for more details on the diagnostic of Markov chain, see [Robert and Casella \(2009\)](#)). In this section, we show how this can be checked using R and the **coda** package. Note that ChronoModel application has also its own diagnostic tools.

In order to use any function from the **coda** package, data has to be transformed in a `mcmc.list` using the function `coda.mcmc()`. This function takes two parameters: the name of the dataset and the number of parallel chains generated by the application.

```
R> Dates_mcmc = coda.mcmc(data = KADatesChronoModel, numberChains = 3)
```

Now, we can trace the history plot of each parameter and see whether the chains have reach their equilibrium before the Acquire period. The column 1 containing the iteration number

should be withdrawn.

```
R> plot(Dates_mcmc[, -1])
```

A part of the results is given in Figure 3. We can see that the 3 chains give similar results. For each parameter, traces visit the same time space of values, traces are correctly overlaid, and they seem to have a stationary behaviour. Indeed, within each subinterval, the same time space is visited. Hence, we can assume that the 3 chains have properly reach the same equilibrium, and finally that the overall MCMC sample (the concatenation of the 3 chains) has converged.

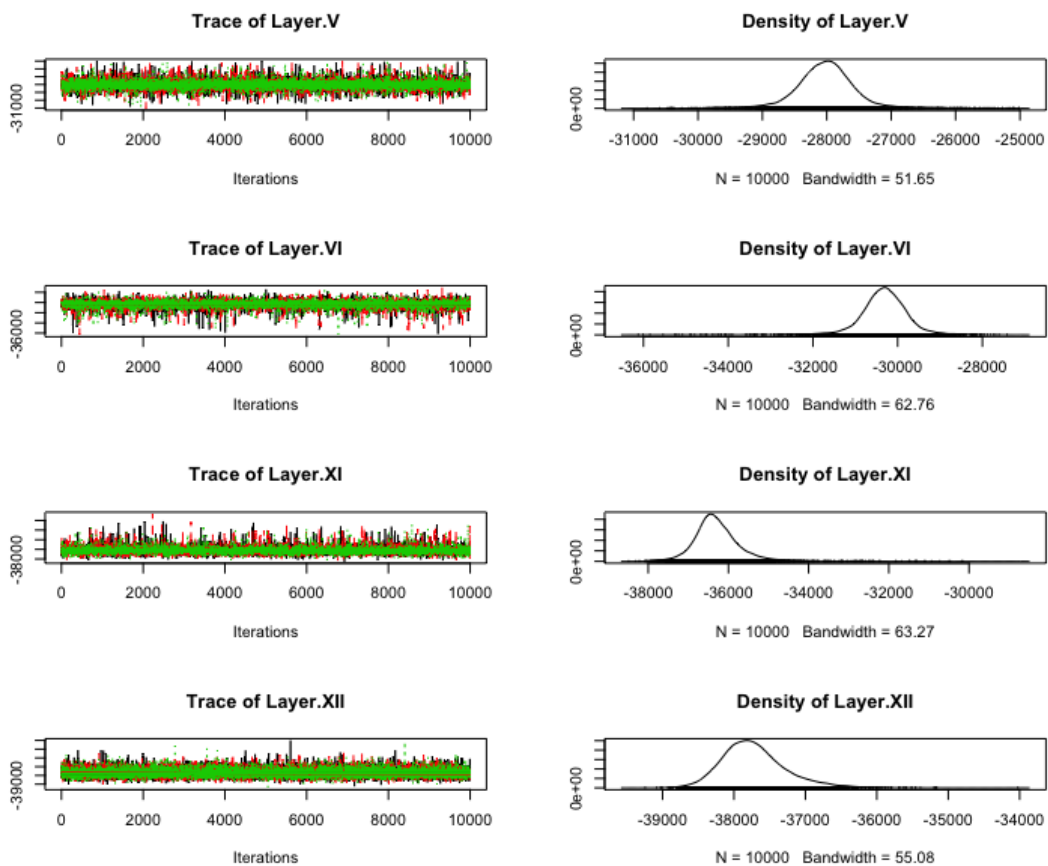


Figure 3: Traces and densities of several parameters of the Bayesian modelling.

The Gelman-Rubin diagnostic can also be tested :

```
R> gelman.diag(Dates_mcmc[, -1])
```

```
> gelman.diag(Dates_mcmc[, -1])
```

Potential scale reduction factors:

Point est. Upper C.I.

Layer.V	1	1
Layer.VI	1	1
Layer.XI	1	1
Layer.XII	1	1
Layer.XVI.4	1	1
Layer.XVI.3	1	1
Layer.XVI.1	1	1
Layer.XVI.2	1	1
Layer.XVII.2	1	1
Layer.XVII.1	1	1
Layer.XVII.3	1	1
Layer.XVII.4	1	1
Layer.XVIII	1	1
Layer.XIX	1	1
Layer.XX	1	1
Layer.XXII	1	1

Multivariate psrf

1

The values of the Gelman-Rubin criterion are equal to 1 meaning that the different Markov chains run in parallel have converged.

In summary, the Markov chains have properly reached their equilibrium. We can concatenate the 3 chains and go further in the analysis.

#### 4.4. Examining a series of dates

First, we analyse the individual marginal posterior density of the 16 events that correspond to evidence of human activity on the site.

In a Bayesian framework, an estimation of a date can be summarised by its confidence interval given at a 95% probability thanks to the marginal posterior density of the related date. This gives an interval of time that has 95% to contain the true date of interest. This interval is not assumed to have equal tail probabilities on the left and right but it is optimised in order to be the shortest. However, this statistical tool is not adapted if the posterior density is bimodal, Highest Posterior Density (HPD) regions would be preferred.

The estimation of credible interval can be done for a collection of dates using the function `MultiCredibleInterval` that takes at least two parameters : `data` the name of the dataset in which the MCMC samples are stocked, `position` the column numbers referring to the MCMC sample of the dates of interest in the dataset.

R> `MultiCredibleInterval(KADatesChronoModel, position = c(2:17), level = 0.95)`  
which gives the following results :

```
> MultiCredibleInterval(KADatesChronoModel, position = c(2:17), level = 0.95)
      Level CredibleIntervalInf CredibleIntervalSup
Layer.V      0.95             -29071             -27105
Layer.VI      0.95             -31492             -29123
```

Layer.XI	0.95	-37422	-35021
Layer.XII	0.95	-38546	-36756
Layer.XVI.4	0.95	-40305	-37791
Layer.XVI.3	0.95	-40340	-37826
Layer.XVI.1	0.95	-39244	-37589
Layer.XVI.2	0.95	-39997	-38349
Layer.XVII.2	0.95	-40821	-38970
Layer.XVII.1	0.95	-40992	-39047
Layer.XVII.3	0.95	-41196	-39566
Layer.XVII.4	0.95	-40793	-39606
Layer.XVIII	0.95	-41417	-40124
Layer.XIX	0.95	-41626	-40511
Layer.XX	0.95	-42130	-40851
Layer.XXII	0.95	-43215	-41105

As expected, the evidence of human activity in Layer XXII, the deepest dated layer, is associated with the oldest 95% confidence interval. Layer XVII and layer XVI are associated with four dates as four shellfish were found in them. The 95% confidence interval of the four evidences of human activity in Layer XVII are more or less similar.

A Highest Posterior Density (HPD) region (see [Hyndman \(2018\)](#)) is another way of summarising the estimation of a parameter in a Bayesian framework. It is more adapted than credible interval in the case of multimodal densities. Otherwise, it gives the same interval as the optimised credible interval calculated by `CredibleInterval()` function. The function `MultiHPD` gives the HPD region of the specified collection of dates (results not shown).

```
R> MultiHPD(KADatesChronoModel, position = c(2:17), level = 0.95)
```

All those intervals may be drawn on a graph using the function `MultiDatesPlot`. The following lines show two examples of its use, but only the result of the first line is presented in [Figure 4](#).

```
R> MultiDatesPlot(KADatesChronoModel, position = c(2:17), level = 0.95,
intervals="CI", title=" 95% CI of the dates", order = "increasing")
R> MultiDatesPlot(KADatesChronoModel, position = c(2:17), level = 0.95,
intervals="HPD", title=" 95% HPD regions of the dates", order = "increasing")
```

We can see that dates are properly sorted according to the stratigraphic order. Confidence intervals are close to each other from layer to layer, except between layer XI and layer VI. Indeed no shellfish were found in the layers VII to X leading to a gap of time in this series of dates.

We can examine the rate of occurrence of the human activity throughout the stratigraphy. To do that two functions may be used : `TempoPlot` and `TempoActivityPlot`.

```
R> TempoPlot(KADatesChronoModel, position = c(2:17), level = 0.95,
title=" Tempo plot")
R> TempoActivityPlot(KADatesChronoModel, position = c(2:17), level = 0.95)
```

[Figure 5](#) displays the tempo plot and the activity plot of the evidences of human activity found

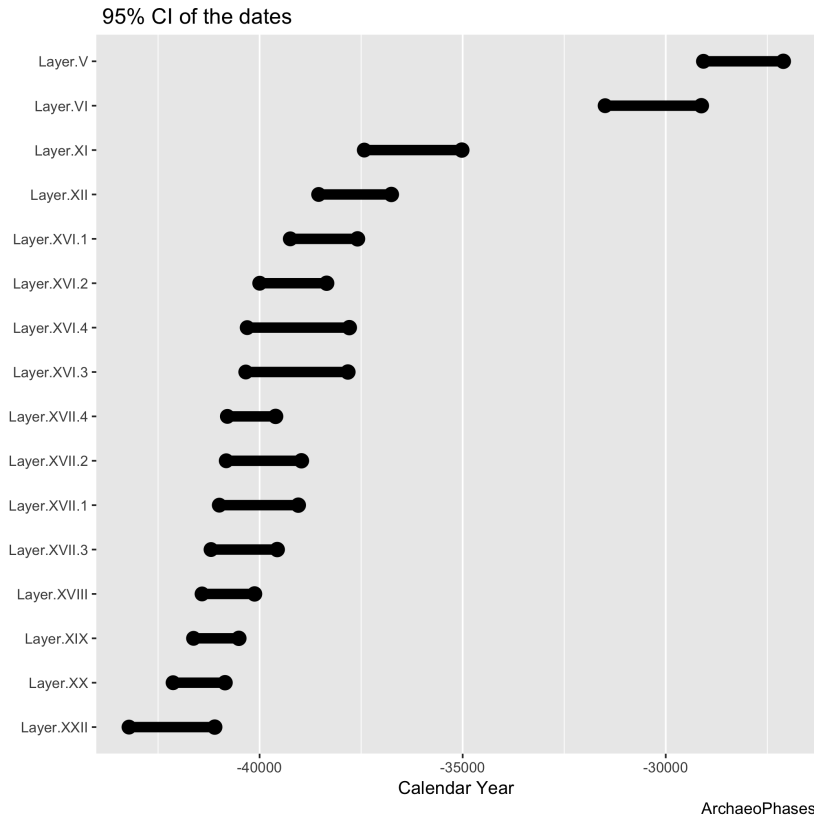


Figure 4: Graphical summary of the estimated dates included in the chronology of Ksâr 'Akil

at Ksâr 'Akil. From these graphs, we can see that the highest part of the human activity is dated between -43 000 to -35 000 and two evidences are younger, at about -31 000 and -28 000. The activity plot shows that there is an increasing amount of activity after -43 000 with a peak at -40 000, then the activity decreases until -35 000 where it is null. Then two other activities appear later in time. There seem to be a gap of time between the first 14 dates and the two remaining.

#### 4.5. First conclusions

From these three last graphs, we can see that there is a continuity in the occurrences of human activity found at Ksâr 'Akil between -43 000 and -35 000. This correspond to the activity found in layers XXII to XI. Then there is a hiatus until -29 000. Then two other occurrences of human activity appears. These occurrences were found in layer VI and layer V. The hiatus highlighted there may correspond to the hiatus linked to the time elapsed between the deposit of layer XI and of layer VI.

In this context, the Tempo Plot and the activity plot reflect more the sampling of the evidences of human activity, that is the sampling of a peculiar kind of consumed shellfish throughout the stratigraphy. In addition, only 16 of these shellfish could be dated by radiocarbon. This is probably due to a random process. In some layers, several shells were found. and dated. Hence human activity related to these layers were more precisely dated than in layer with only one shell. The gap of time is probably due to the absence of dated shells rather than to



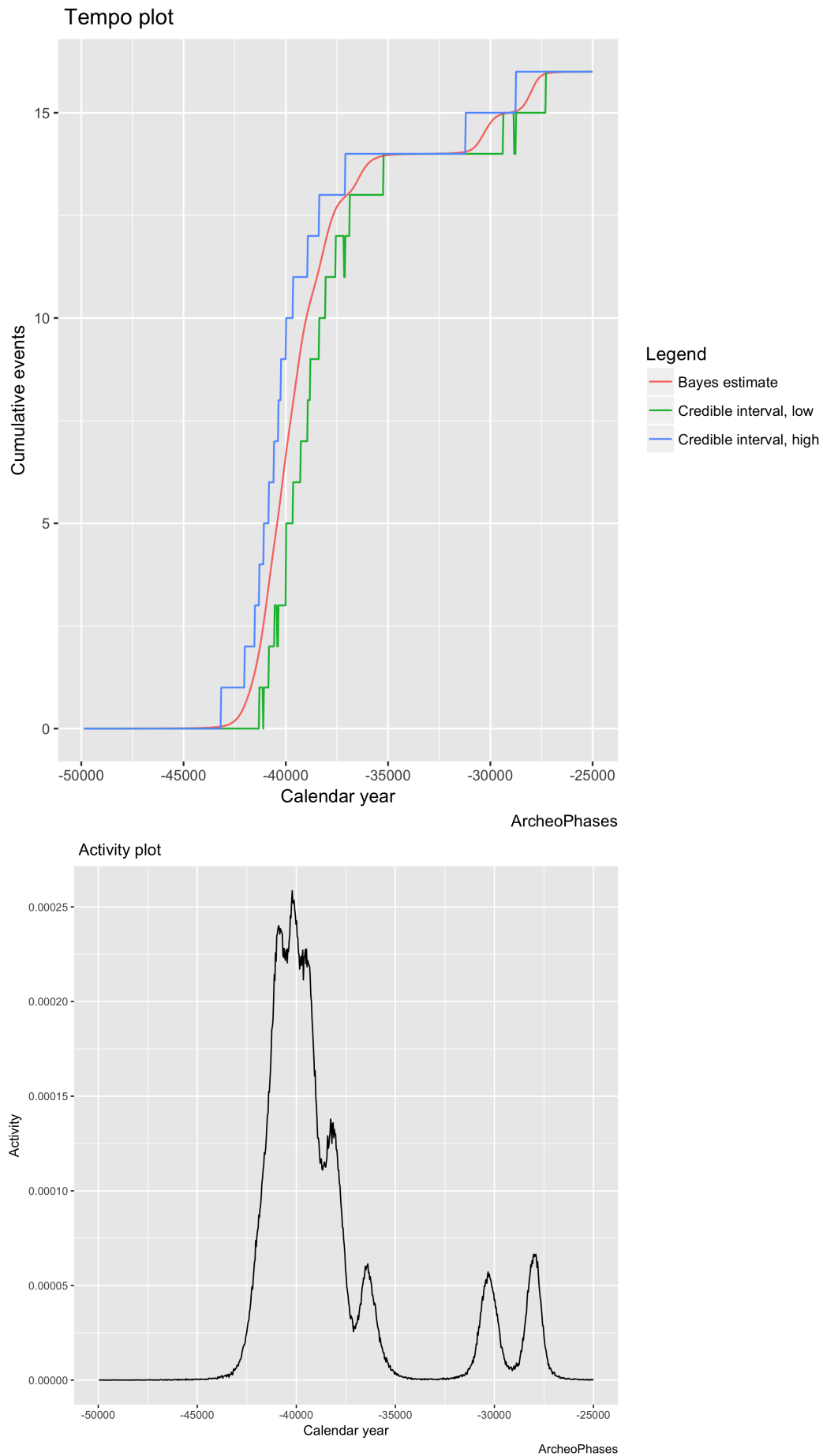


Figure 5: Tempo plot [top] and activity plot [bottom] of the chronology of Ksâr 'Akil

the absence of human activity.

In all, human activity found at Ksâr 'Akil is dated at a period of time in coherence with other results, such as Bosch *et al.* (2015); Douka, Bergman, Hedges, Wesselingh, and Higham (2013). However, the analysis can be further refined by the inspection of the Palaeolithic periods of time, called *phases* hereafter, and by the use of our new statistical tools.

## 5. Palaeolithic chronology

The lithic assemblages found throughout the stratigraphy of Ksâr 'Akil allowed us to associate the first XXV layers to four Palaeolithic periods: Initial Upper Palaeolithic (IUP), from layer XXV (the bottom of the sequence) to layer XXI; Ahmarian, from layer XX to layer XIV; Upper Palaeolithic (UP), from layer XV to layer VI; and Epi-Palaeolithic (EPI), from layer V to layer I. The aim of this modelling is to establish the chronology of these four successive Palaeolithic periods thanks to the chronology of the stratigraphy.

In **ChronoModel**, we defined 4 *phases* in order to estimate the characteristics of these four Palaeolithic periods. As said in Section 3, several CSV files are created by **ChronoModel** and in particular, a file called "phases.csv" containing an estimation of the minimum date (called alpha) and maximum date (called beta) of each *phase*, as long as at least one *phase* is modelled. In **ChronoModel** the start date (resp. end date) of a *phase* is estimated by the minimum date (resp. the maximum date) of the collection of dates associated to this *phase*.

In the **ArchaeoPhases** package, a dataset containing the MCMC samples of the start date (called alpha) and the end date (called beta) of each Palaeolithic *phase*, and generated by **ChronoModel** application, is already included.

```
R> data(KAPhasesChronoModel)
R> attach(KAPhasesChronoModel)
```

Let us first analyse a unique period of time, the period called Ahmarian, then we will examine the succession of the four Palaeolithic periods.

### 5.1. Examining the Ahmarian period

Let us analyse the Ahmarian period that gathers 11 dates of the 16 from the chronology.

#### *Classical summary*

A period of time is classically summarised by its start and end dates. From these two dates, we can estimate the duration of the period. The function **PhaseStatistics()** gives the classical summary statistics of the characteristics of a *phase*: mainly mean, median, credible interval and HPD region. This function takes a minimum of two parameters: a MCMC sample of the marginal posterior distribution of the start date and a MCMC sample of the marginal posterior distribution of the end date.

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

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

```
> PhaseStatistics(Ahmarian.alpha, Ahmarian.beta, level = 0.95)
      Minimum      Maximum Duration
```

mean	-41463.00	-38249.00	3214.00
MAP	-41436.00	-38248.00	3207.00
sd	334.00	403.00	517.00
Q1	-41659.00	-38496.00	2875.00
median	-41437.00	-38249.00	3202.00
Q2	-41238.00	-38001.00	3535.00
level	0.95	0.95	0.95
CredibleInterval Inf	-42130.00	-39076.00	2200.00
CredibleInterval Sup	-40851.00	-37456.00	4242.00
HPDRInf 1	-42092.00	-39063.00	2183.00
HPRDSup 1	-40801.00	-37437.00	4227.00

`PhaseStatistics()` results in a matrix of several summary statistics according to the marginal posterior density of the start date, the end date and the duration of the *phase*. The default confidence level is 0.95.

The start of Ahmarian is estimated between -42 130 and -40 851 (dates are in calendar year Before Christ) with a 95% probability. Its end date is estimated between -39 076 and -37 456, and its duration between 2 200 to 4 242 years.

### *Time Range interval*

We can estimate locate in time the Ahmarian period by its time range interval. This gives an interval of time that resumes the start, the end and the duration of the period at a given confidence level.

The following code gives the endpoints of the time range interval of the *phase* Ahmarian.

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

The output of the console is

```
> PhaseTimeRange(Ahmarian.alpha, Ahmarian.beta, level = 0.95)
      level TimeRangeInf TimeRangeSup
      0.95    -42189.00    -37461.00
```

Hence, with a 95% probability, the period Ahmarian started at -42 189 and ended at -37 461 (dates are in calendar year Before Christ). And this Palaeolithic period lasted 4 728.

The endpoints of the time range of Ahmarian are close to the inferior endpoint of the 95% credible interval of the start date and the superior endpoint of the 95% credible interval of the end date. However, only the time range gives an interval that covers the collection of dates of the *phase* with a 95% probability.

### *Graphical illustration*

Function `PhasePlot()` may be used to draw a plot of the marginal posterior density of the start and end dates of a *phase* on the same graph. 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 result is shown on Figure 6 and displays the characteristics of a *phase*.

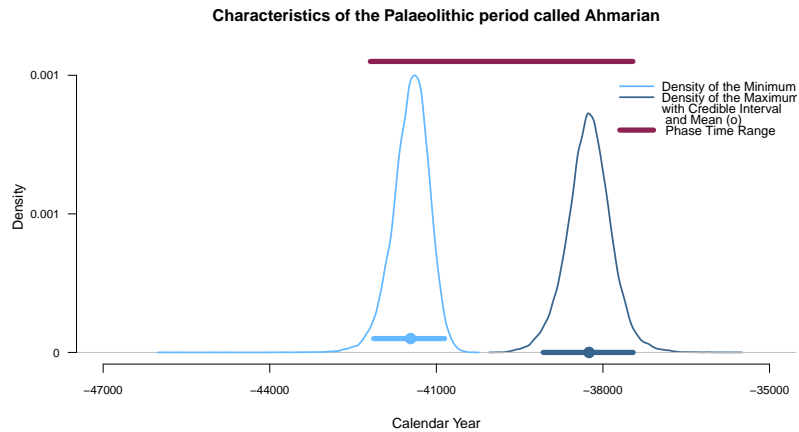


Figure 6: Characteristics of the Palaeolithic period called "Ahmarian". The density of the start date is drawn in dashed line, the density of the end date is drawn in solid line. The segment above the curve represents the 95% time range interval.

```
R> PhasePlot(Ahmarian.alpha, Ahmarian.beta, level = 0.95,
title ="Characteristics of the Palaeolithic period called Ahmarian")
```

## 5.2. Examining the four Palaeolithic periods

Now, let us analyse the succession of the four Palaeolithic periods: IUP, Ahmarian, UP and EPI.

### *Time range intervals*

The function `MultiPhaseTimeRange()` gives the time range intervals of several *phases*. It takes at least two parameters, `data` the name of the dataset, and `position_minimum` is a vector of positions that identify the start dates of the corresponding *phases* in the data matrix. By default, the vector of positions that identify the end dates is assumed to be `position_maximum = position_minimum + 1`.

Let's estimate the time range of phases IUP, Ahmarian, UP and EPI, whose start date outputs are respectively in column number 8, 6, 4 and 2 in the data matrix.

```
R> MultiPhaseTimeRange(data=KAPhasesChronoModel, position_minimum = c(8,6,4,2))
```

```
> MultiPhaseTimeRange(data = KAPhasesChronoModel, position_minimum = c(8,6,4,2))
      Level TimeRangeInf TimeRangeSup
IUP.alpha IUP.beta      0.95      -43217      -41106
Ahmarian.alpha Ahmarian.beta 0.95      -42189      -37461
UP.alpha UP.beta      0.95      -38559      -29335
EPI.alpha EPI.beta      0.95      -29071      -27102
```

Phase "EPI" whose start date output is in column 2 has a time range = [-29 071, -27 102] at a confidence level of 95% and the phase "IUP" whose start date output is in column 8 has a time range = [-43 217, -41 106].

*Transition intervals*

Now, let us estimate the transition intervals between two successive *phases*.

```
R> MultiPhasesTransition(KAPhasesChronoModel, position_minimum=c(8,6,4,2))
```

For this functions the order of the succession is important. The vector of positions of the start date (referring to the column number in the data matrix) should be ordered from the oldest *phase* to the youngest one.

```
> MultiPhasesTransition(KAPhasesChronoModel, position_minimum=c(8,6,4,2))
              0.95 TransitionRangeInf TransitionRangeSup
IUP.beta & Ahmarian.alpha 0.95          -43241          -40728
Ahmarian.beta & UP.alpha  0.95          -39075          -36686
UP.beta & EPI.alpha       0.95          -31504          -26960
```

The transition period between IUP and Ahmarian is estimated from -43 241 (in calendar years Before Christ) to -40 728. Between Ahmaria and UP the transition period started at -39 075 till -36 686. Between UP and EPI the transition period is estimated from -31 504 to -26 960.

*Gap intervals*

We can test if a the gap between these successive *phases* exists. This may be done using the following code :

```
R> MultiPhasesGap(KAPhasesChronoModel, position_minimum=c(8,6,4,2))
```

Again, for this function, the order of the succession is important. The vector of the start position (referring to the column number in the data matrix) should be ordered from the oldest *phase* to the youngest one.

```
> MultiPhasesGap(KAPhasesChronoModel, position_minimum=c(8,6,4,2))
              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          -29180        -28977
```

With probability 95% there is no gap between the succession of phases IUP, Ahmarian and UP. Indeed we cannot construct an interval satisfying (5) with posterior probability 95% . On the other hand there exists a gap of 203 years between phase UP and phase EPI. This gap interval, with posterior probabability 95% , starts at -29 180 and ends at -28 977.

*Graphical illustration*

The function `MultiSuccessionPlot()` draws the characteristics of a succession of groups. Figure 7 presents the resulting plot.

```
R> MultiSuccessionPlot(KAPhasesChronoModel, position_minimum =c(8,6,4,2),
title = "Ksar Akil - Succession of phases: IUP, Ahmarian, UP, EPI ")
```

In this plot, a group of dates is identified by a color.

- The density curves represent the densities of the minimum (solid line) or the maximum (dashed line) 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 (see the phase IUP in Figure 7).
- Segments above the density curves correspond to time range of the phase.
- Two-coloured segments correspond to transition interval and to the gap range. When there is no gap between the succession of phases, the two-coloured segment is replaced by a cross (see Figure 7 between phases IUP and Ahmarian, and Ahmarian and UP).

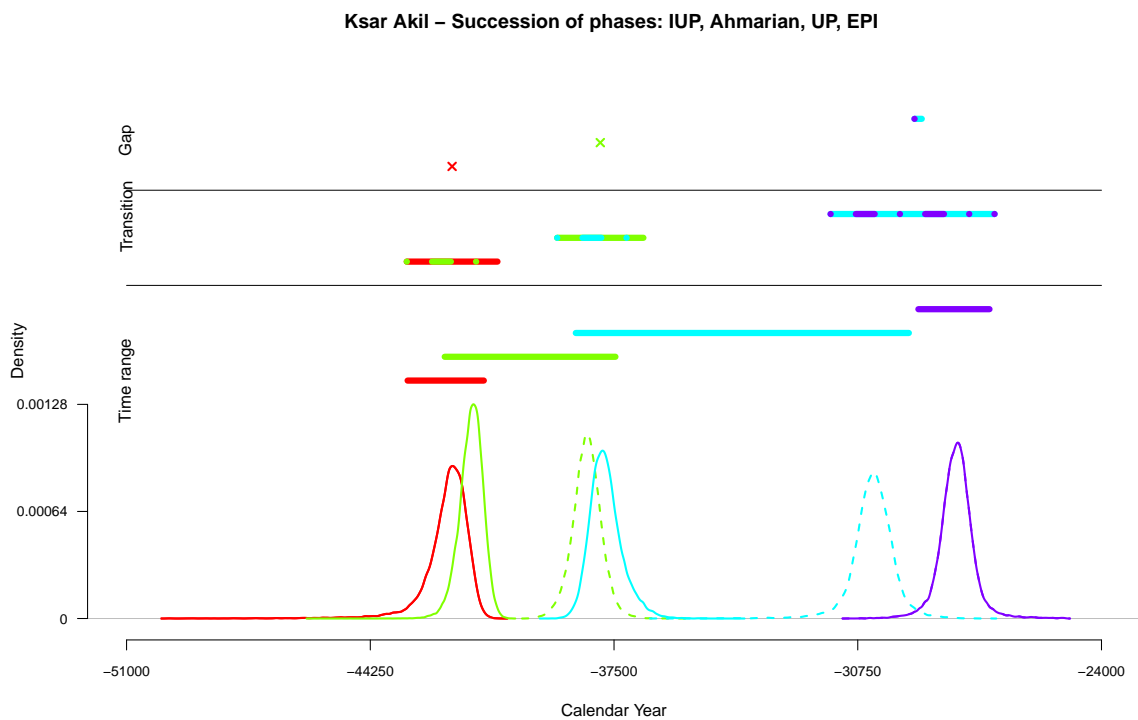


Figure 7: Plot of the succession of phases from Ksâr 'Akil. The characteristics of phase IUP are drawn in green, those of phase Ahmarian are in purple, those of phase UP are in light blue and those of phase EPI are in red.. Time range are displayed by segments above the curves. Two-coloured segments correspond to transition interval or to the gap range. As there are no gaps between phases IUP and Ahmarian, and Ahmarian and UP, a cross is drawn instead.

### 5.3. Summary

We can summary the succession of the four Palaeolithic periods estimated from the site of Ksâr 'Akil, using the data published by Bosch *et al.* (2015) and the modeling done with ChronoModel described in section 3.2, by the following tables :

Results from Table 1 are similar as those given by Bosch *et al.* (2015). We can see that for phases containing only one events, such as EPI and IUP, the 95% credible interval of the start

Palaeolithic period	Start date		End date		Time range	
	Inf	Sup	Inf	Sup	Start	End
<b>IUP</b>	-43 215	-41 105	-43 215	-41 105	-43 217	-41 106
<b>Ahmarian</b>	-42 130	-40 851	-39 076	-37 456	-42 189	-37 461
<b>UP</b>	-38 546	-36 756	-31 492	-29 123	-38 559	-29 335
<b>EPI</b>	-29 071	-27 105	-29 071	-27 105	-29 071	-27 102

Table 1: Classical and new summaries of the four Palaeolithic *phases* in succession. Endpoints (Inf and Sup values) of 95% credible interval of the start and end dates of each *phase* and 95% Time range intervals. Dates are in years BCE.

Succession of Palaeolithic periods	Transition range		Gap range	
	Start	End	Start	End
<b>IUP &amp; Ahmarian</b>	-43 241	-40 728	<b>No Gap</b>	
<b>Ahmarian &amp; UP</b>	-39 075	-36 686	<b>No Gap</b>	
<b>UP &amp; EPI</b>	-31 504	-26 960	-29 180	-28 977

Table 2: New summary statistics of the four Palaeolithic *phases* in succession : transition range intervals and gap range intervals given with a level of confidence at 95%. Dates are in years BCE.

(or end) date is equal (up to numeric approximations) to the 95% time range interval. For *phases* containing several events such as Ahmarian and UP the time range interval provide rigorous alternative to the usual approach consisting to take the interval whose endpoints are the inferior value of the credible interval of the start date and the superior value of the credible interval of the end date. However, only the time range interval gives a 95% interval for the period covering all the dates included in the *phase*.

Table 2 gives also the endpoints of the transition range intervals and the gap range intervals if existing. Again, the 95% transition intervals are similar to the interval build with the inferior value of the 95% credible interval of the end date of the older *phase* and the superior value of the 95% credible interval of the start date of the younger *phase*. However, the only interval that gives a 95% probability to include all the dates is the transition interval.

The gap range intervals are similar to an interval build with the superior value of the 95% credible interval of the end date of the older *phase* and the inferior value of the 95% credible interval of the start date of the younger *phase*.

Our new statistical tools aim at summarizing periods of time by intervals that cover all the dates with a fixed probability. Instead of summarizing a *phase* by two credible intervals, one for the start date, one for the end date, the time range interval is a unique interval given with fixed probability.

In conclusion, we can summarize the succession of the Paleolithic phases by saying that from the data found at from the site of Ksâr 'Akil, using the data published by Bosch *et al.* (2015) and the modeling done with ChronoModel described in section 3.2, with a 95% probability, Phase IUP is estimated from -43 217 BCE to -41 106 BCE, Phase Ahmarian is estimated from -42 189 to -37 461, Phase UP is estimated from -38 559 to -29 335 and Phase EPI is estimated from -29 071 to -27 102 (years in BCE). The transition periods between successive

phases are estimated with a 95% probability.

## 6. Conclusion

The **ArchaeoPhases** package provides new statistical tools that help interpret the MCMC sample generated from the posterior distribution of a chronology of dates. Among these tools, two of them explore the rate of occurrence of dates (`TempoPlot()` et `TempoActivityPlot()`). Three other tools summarise periods of time, or what we call *phases*, defined by some archaeological, paleo-environmental, geographical... criterion. A *phase* is associated with a collection of dates estimated from the chronology. `PhaseTimeRange()` (or `MultiPhaseTimeRange()`) gives an interval of time that characterizes the period corresponding to that *phase* with a fixed probability. `PhasesTransition()` (or `MultiPhasesTransition()`) gives an interval of time that characterizes the transition between two successive *phases* with a fixed probability. `PhasesGap()` (or `MultiPhasesGap()`) tests the existence of a hiatus between two successive *phases*, and if so, gives an interval of time that characterizes the hiatus between these two successive *phases* with a fixed probability. See Guérin, Antoine, Schmidt, Goval, Hérissou, Jamet, Reyss, Shao, Philippe, Vibet, and Bahain (2017) and Philippe and Vibet (2018) for an application of these tools.

We chose to illustrate the use of these statistical tools, in comparison with the classical ones, by the analysis of four Palaeolithic periods and the stratigraphy of Ksâr 'Akil. The Bayesian modelling made with ChronoModel, a software application adapted for the construction of chronologies of estimated dates, resulted in a chronology similar to the one given by Bosch *et al.* (2015). Our new tools add complementary information to classical ones. The time range interval summarises a *phase* by a unique interval with a fixed probability. Transition range and gap range intervals result in a unique interval that defines a transition between two *phases* or a gap between two *phases*.

The **ArchaeoPhases** package has been applied to archaeology (see Philippe and Vibet (2018); Duvivier and Florent (2018); Philippe, Vibet, and Bahain (2018)) but it could also be used in other fields such as paleo-environnement (see Guérin *et al.* (2017)), forensic physics and so on as long as dates are estimated.

## Acknowledgments

Marie-Anne Vibet thanks the DefiMaths Program and Region Pays de La Loire for financial supports. Correspondence should be sent to Anne Philippe (anne.philippe@univ-nantes.fr).

## 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 'Akil (Lebanon) Supports Levantine Route of Modern Human Dispersal into Europe." *Proceedings of the National Academy of Sciences*, **112**(25), 7683–7688. doi:10.1073/pnas.1501529112.



- Bronk Ramsey C (2009). “Bayesian Analysis of Radiocarbon Dates.” *Radiocarbon*, **51**(1), 337–360.
- Bronk Ramsey C (2016). *Oxcal 4.2*. URL [http://c14.arch.ox.ac.uk/oxcalhelp/hlp\\_contents.html](http://c14.arch.ox.ac.uk/oxcalhelp/hlp_contents.html).
- Buck C, Boden G, Christen A, James G, Sonnenwald F (2018). *BCal*. URL <https://bcal.sheffield.ac.uk>.
- Buck C, Christen A, James G (1999). “BCal: an on-line Bayesian radiocarbon calibration tool.” *Internet Archaeology*, **7**. URL <https://doi.org/10.11141/ia.7.1>.
- Buck C, Litton C, Smith A (1992). “Calibration of radiocarbon results pertaining to related archaeological events.” *Journal of archaeological Science*, **19**, 497–512.
- Buck CE, Litton CD, Cavanagh WG (1996). *The Bayesian Approach to Interpreting Archaeological Data*. Chichester, J.Wiley and Son, England.
- Christen JA (1994). “Summarizing a set of radiocarbon determinations: a robust approach.” *Applied Statistics*, **43**(3), 489–503.
- Dean JS (1978). “Independent Dating in Archaeological Analysis.” *Advances in Archaeological Method and Theory*, **1**, 223–255. ISSN 01628003. URL <http://www.jstor.org/stable/20170134>.
- Douka K, Bergman CA, Hedges REM, Wesselingh FP, Higham TFG (2013). “Chronology of Ksar Akil (Lebanon) and implications for the colonization of Europe by anatomically modern humans.” *PloS one*, **8**(9), e72931. ISSN 1932-6203. doi:10.1371/journal.pone.0072931. URL <http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0072931>{#}pone.0072931.s008.
- Duvivier H, Florent G (2018). “Un établissement agricole du Premier Age du Fer à La Chapelle d’Armentieres ”Route nationale” (Nord) :un exemple d’habitat á enclos.” *La revue du Nord*. To appear.
- 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.
- Guérin G, Antoine P, Schmidt E, Goval E, Hérison D, Jamet G, Reyss JL, Shao Q, Philippe A, Vibet MA, Bahain JJ (2017). “Chronology of the Upper Pleistocene loess sequence of Havrincourt (France) and associated Palaeolithic occupations: A Bayesian approach from pedomorphology, OSL, radiocarbon, TL and ESR/U-series data.” *Quaternary Geochronology*, **42**, 15–30. ISSN 18711014. doi:10.1016/j.quageo.2017.07.001. URL <http://linkinghub.elsevier.com/retrieve/pii/S1871101416301819>.
- Hyndman RJ (2018). *hdrcde: Highest Density Regions and Conditional Density Estimation*. R package version 3.2, URL <http://pkg.robjhyndman.com/hdrcde>.
- Lanos P, Philippe A (2017). “Hierarchical Bayesian Modeling for Combining Dates in Archaeological Context.” *Journal de la SFdS*, **158**, 72–88.

- Lanos P, Philippe A (2018). “Event date model: a robust Bayesian tool for chronology building.” *Communications for Statistical Applications and Methods*, **25**, 131 – 157.
- Lanos P, Philippe A, Lanos H, Dufresne P (2016). *Chronomodel : Chronological Modelling of Archaeological Data using Bayesian Statistics (Version 1.5)*. URL <http://www.chromodel.fr>.
- Naylor JC, Smith AFM (1988). “An Archaeological Inference Problem.” *Journal of the American Statistical Association*, **83**(403), 588–595.
- Nicholls G, Jones M (2002). “New Radiocarbon Calibration Software.” *Radiocarbon*, **44**, 663–674.
- Philippe A, Vibet MA (2017). *ArchaeoChron: Bayesian Modeling of Archaeological Chronologies*. R package version 0.1, URL <https://CRAN.R-project.org/package=ArchaeoChron>.
- Philippe A, Vibet MA (2018). “A Note on the Bayesian Modeling of the Stratigraphic Chronology of Canimar Abajo, Cuba.” *Radiocarbon*, pp. 1–11. doi:10.1017/RDC.2018.19.
- Philippe A, Vibet MA, Bahain JJ (2018). “Chronologie de mise en place des différentes couches stratigraphiques du site d’Havrincourt.” *Les Occupations moustériennes et Gravettiennes d’Havrincourt*. To appear.
- Robert C, Casella G (2009). *Introducing Monte Carlo Methods with R*. Springer Science & Business Media.
- Vibet MA, Philippe A, Lanos P, Dufresne P (2016). *ChronoModel V1.5 User’s Manual*. URL <http://www.chromodel.fr>.

**Affiliation:**

Anne Philippe

Université de Nantes, Laboratoire de mathématiques Jean Leray

2, Rue de la Houssinière, 44000 Nantes, FRANCE

E-mail: [anne.philippe@univ-nantes.fr](mailto:anne.philippe@univ-nantes.fr)

URL: <http://www.math.sciences.univ-nantes.fr/~philippe/info.html>

Marie-Anne Vibet

Université de Nantes, Laboratoire de mathématiques Jean Leray

2, Rue de la Houssinière, 44000 Nantes, FRANCE

E-mail: [marie-anne.vibet@univ-nantes.fr](mailto:marie-anne.vibet@univ-nantes.fr)