

ChronoModel

modélisation chronologique des données archéologiques avec la statistique bayésienne



Philippe Lanos & Philippe Dufresne

Equipe chronologie-référentiels, UMR 5060 IRAMAT, CRP2A, Université Bordeaux-Montaigne

Equipe Paleo2D, UMR 6118 Géosciences-Rennes, Université de Rennes 1

Les journées de l'agence Lebesgue
Mathématiques et Environnement

Jeudi 11 octobre 2018

Rennes, Campus de Beaulieu, Bât 14B, salle de conférences



Une expérience interdisciplinaire entre statistique et archéométrie

Une collaboration née d'un besoin de mieux formaliser le calcul statistique des dates, suite aux :

- Travaux précurseurs menés sur le 14C et la TL/OSL à partir des années 90
- Travaux menés en archéomagnétisme à Rennes à partir des années 2000

Besoin de calculer des intervalles de confiance qui tiennent compte des erreurs dans le processus de calcul des dates: hiérarchie des erreurs

Besoin de combiner de l'information quantitative avec de l'information qualitative (datation absolue et datation relative)

Besoin de combiner des datations venant de méthodes très différentes: 14C, AM, TL/OSL, DC, Archéologie, Histoire....

Besoin de définir un modèle statistique qui suit au mieux le processus de construction d'une chronologie: combinaison de dates, événements, phases, TPQ, TAQ, relations stratigraphiques, relations de succession, information de durée...

Des compétences multidisciplinaires...

Réunion de compétences qui vont des **mathématiques** et de la **Statistique** à la **physique** et à l'**archéologie**,

qui mobilise un ensemble de chercheurs **spécialistes des méthodes de datation** ainsi que des connaissances dans le domaine de l'**informatique** et de la **simulation numérique** (calcul MCMC).

Les algorithmes développés dans le projet ChronoModel font appel à la **statistique bayésienne**, qui a été mise en œuvre selon les deux niveaux d'analyse suivants :

- Analyse des modèles chronologiques d'Evénement et de Phase.
- Analyse hiérarchique des mesures en datation par luminescence.

Le projet a eu pour principal objectif la réalisation de l'application logicielle **ChronoModel**.

Le développement informatique a été confié à une société de service (iFactory) sur la période 2013-2014 et prise en charge par Ph. Dufresne à partir de 2015.

Programmation effectuée en C++ dans un environnement Qt.

Le projet ChronoModel (2011-2018): un réseau de collaborations

Projet ANR ChronoModel (2011-2015)

Modélisation chronologique et développement d'un logiciel pour le traitement des données de datation en archéométrie, archéologie et géosciences.

Partenariat entre Bordeaux, Nantes et Rennes.

Soutien du Centre Henri-Lebesgue (2014-2016)

Poste IR recruté pour valoriser les méthodes statistiques bayésiennes et les outils de traitement de données mettant en jeu le temps (manuel et contrib. version 1.5).

Soutien du Labex LaScArBx Bordeaux (2015-2016)

Post-Doc pour développer de nouveaux outils statistiques pour la datation OSL, en prolongement des travaux menés dans le cadre de l'ANR Chronomodel (BaSAR).

Développement de la version CM 2.0 (2016-2018)

Amélioration des performances du logiciel et de l'ergonomie.

Ajout de nouvelles fonctionnalités, en collaboration avec les utilisateurs.

Diffusion des connaissances

Logiciels produits: ChronoModel / site WEB « chronomodel.fr » / BaSAR

Publications

- 5 sur la modélisation chronologique des « Faits » et des « Phases »
- 3 sur la modélisation des mesures physiques en datation OSL

Formations organisées entre 2010 et 2018

RenDateModel et OxCal :

Bordeaux 2010 ; Saint-Malo (2011), Dinan (2012)
Toulouse (2013), Aix-en-Provence (2013), Toulouse (2014),

Chronomodel: ANR Roscoff et ANF Bibracte (2015)

ANF Roscoff et INRAP Villeneuve-d'Ascq (2018)

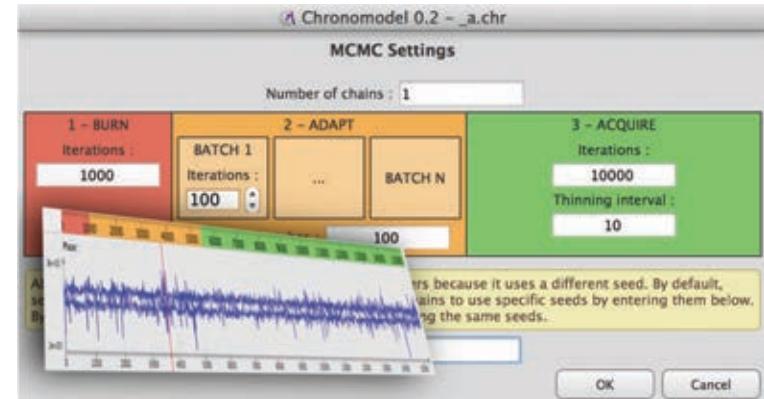
Soutenues par la DR15 CNRS Aquitaine-Limousin, la DR14 CNRS Midi-Pyrénées, l'Universités de Toulouse 2 Le Mirail, l'Université de Aix-Marseille, l'ANR, le réseau « Archéométrie » CAI-RN de la Mission pour l'Interdisciplinarité du CNRS.

Organisation d'un colloque en 2016

Dans le cadre du Semestre Statistique du Centre Henri Lebesgue, 24-26 mai 2016
« Bayesian Statistics applied to Archaeology »

ChronoModel (2015-2018)

<https://chronomodel.com/>



ChronoModel is a free and open-source cross-platform software (Mac, Windows, Linux).

Uses numerical methods based on Markov Chain Monte Carlo (MCMC).

Lanos Ph. and Philippe A., 2017. Hierarchical Bayesian modeling for combining dates in archaeological context. *Journal de la Société Française de Statistique*, Vol. 158 (2) p. 72-88.

Lanos Ph. And Philippe A., 2018. Event date model: a robust Bayesian tool for chronology building. *Communications for Statistical Applications and Methods* (CSAM), Vol. 25, No. 2, 131–157.

Chronology building in archaeology

mobilizes 2 steps which can be formalized in a Bayesian statistical framework

1. Prior information

- We observe stratigraphic relationships during excavations
- We infer some information about the study period, the maximal duration of a group of events, or the minimal hiatus between two groups of events.
- We classify (seriation) groups of artefacts according to “facies”: ceramic, lithic, etc.
- We compare the artefacts with well-dated references, corpus: chronological markers

2. Data acquisition and Model

- We obtain dating determinations from the laboratories : ^{14}C , AM, TL/OSL, DC, etc.
- On utilise le **modèle d'Event** pour relier les mesures aux dates cibles.

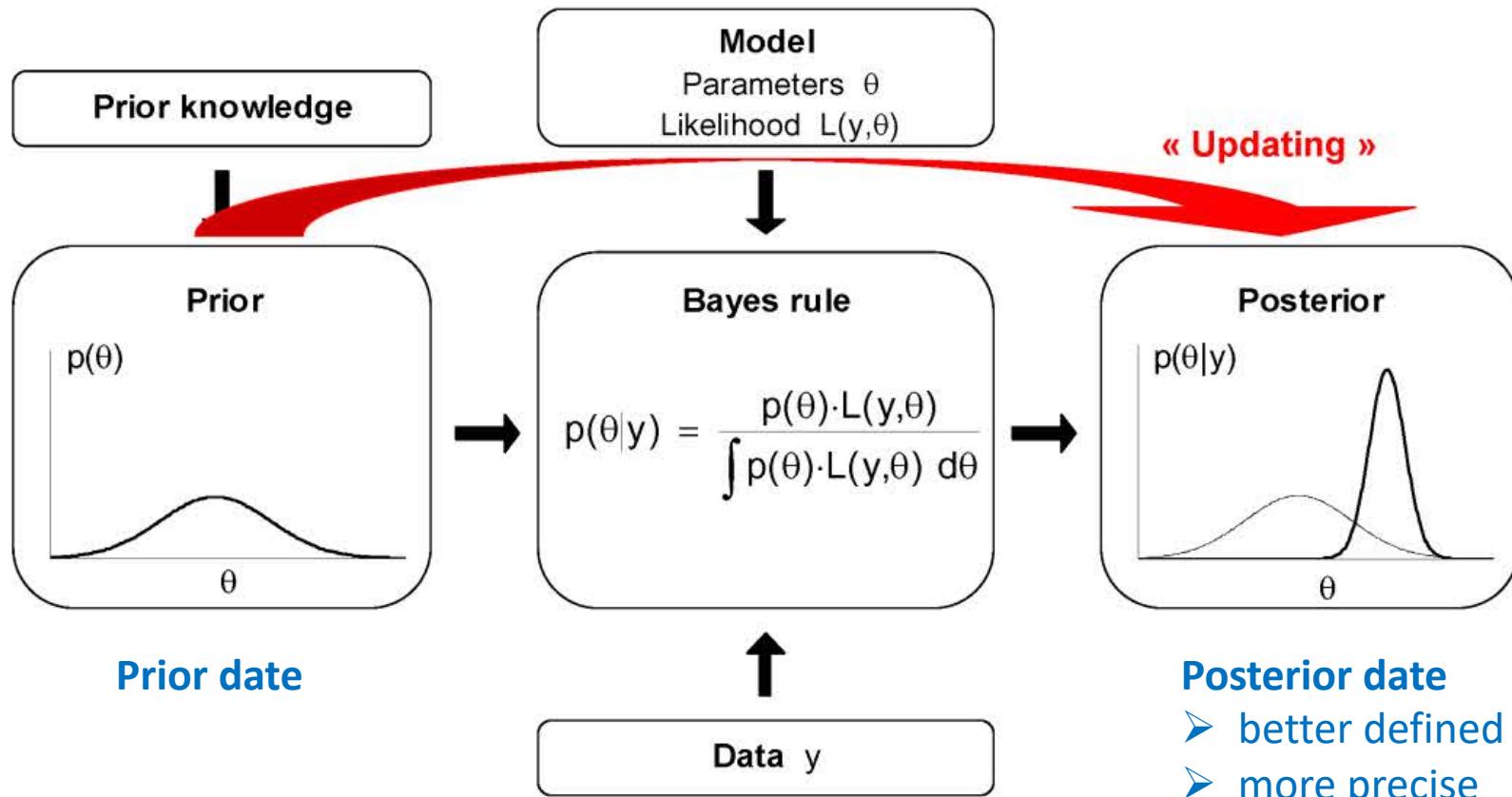
-> Posterior information

- Chronology is obtained as a posterior result thanks to the **Bayesian statistics**.
- We seek for the range of a phase (group of events): **start, end, duration**.
- We aim to estimate the events rhythm : **Tempo-Plot and Activity curve**.



Thomas Bayes (1702–1761)

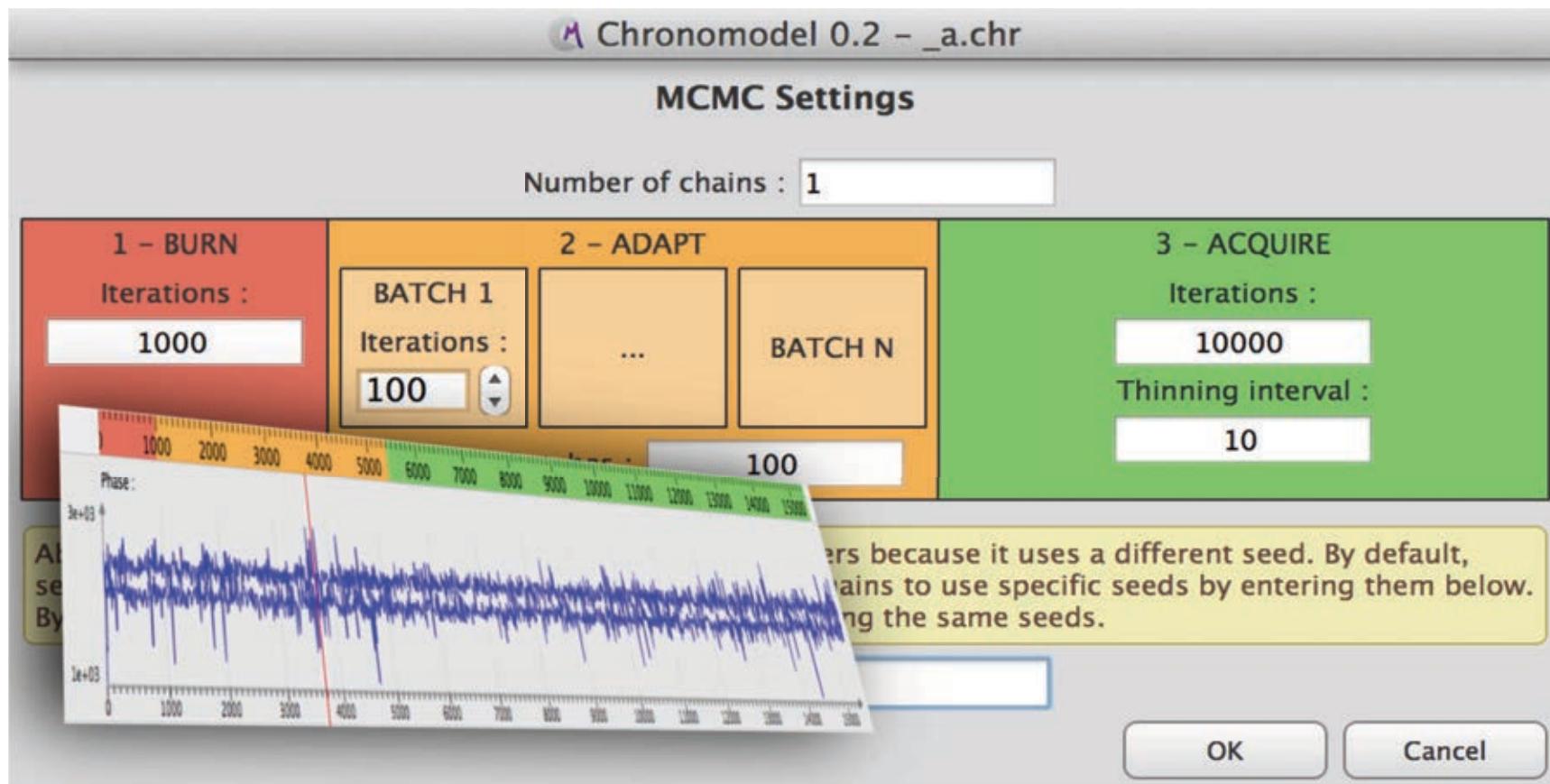
Bayes formula : an information process



Measurement in dating Lab.

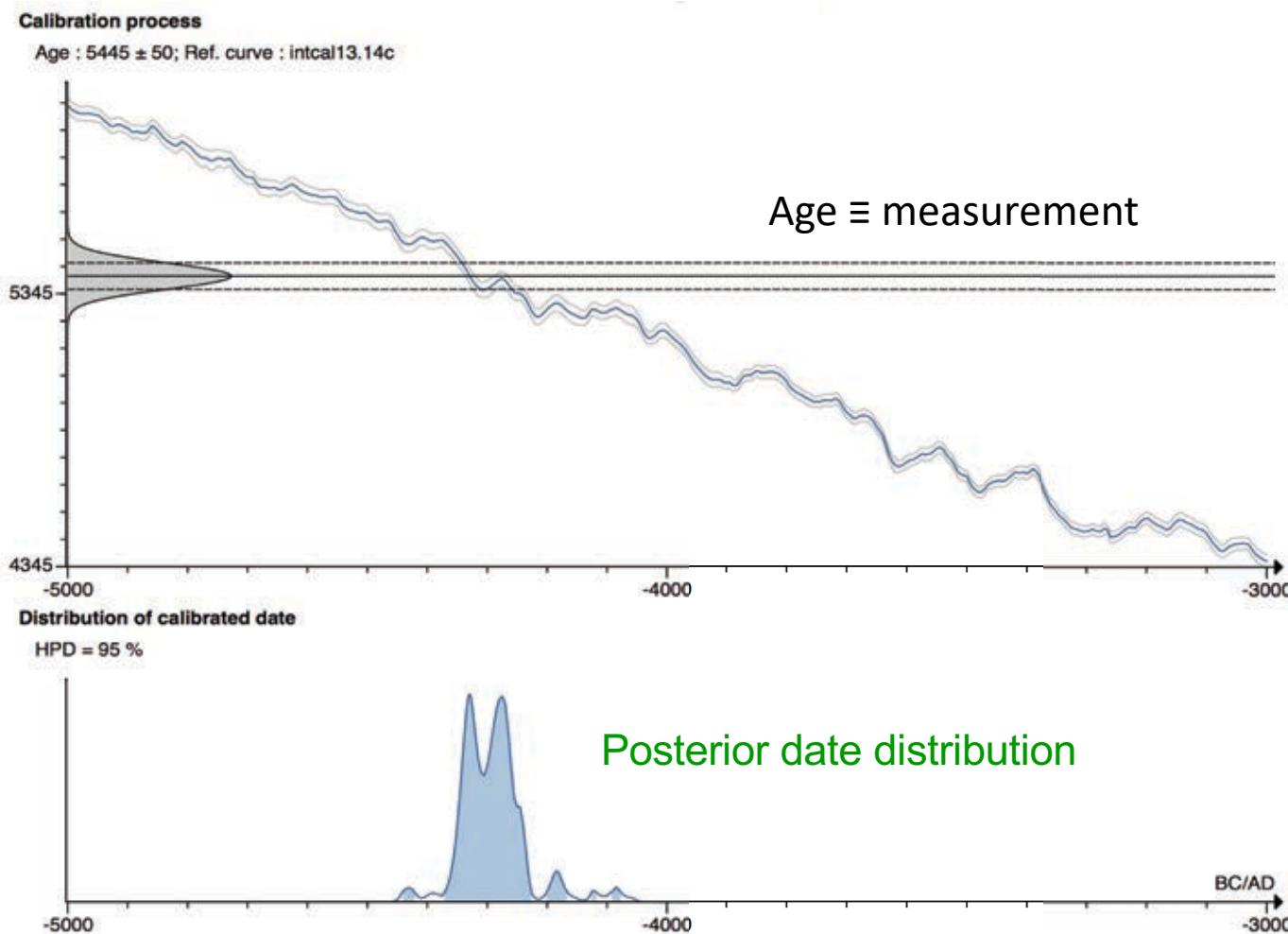
Contrôle des paramètres du MCMC

- Calculation steps explained.
- Personalized seeds & chains length.
- Custom chains number.
- MCMC results analysis : trace, acceptation rate, auto-correlation...
- MCMC per-variable options : different methods for events & data.
- Initialization log.



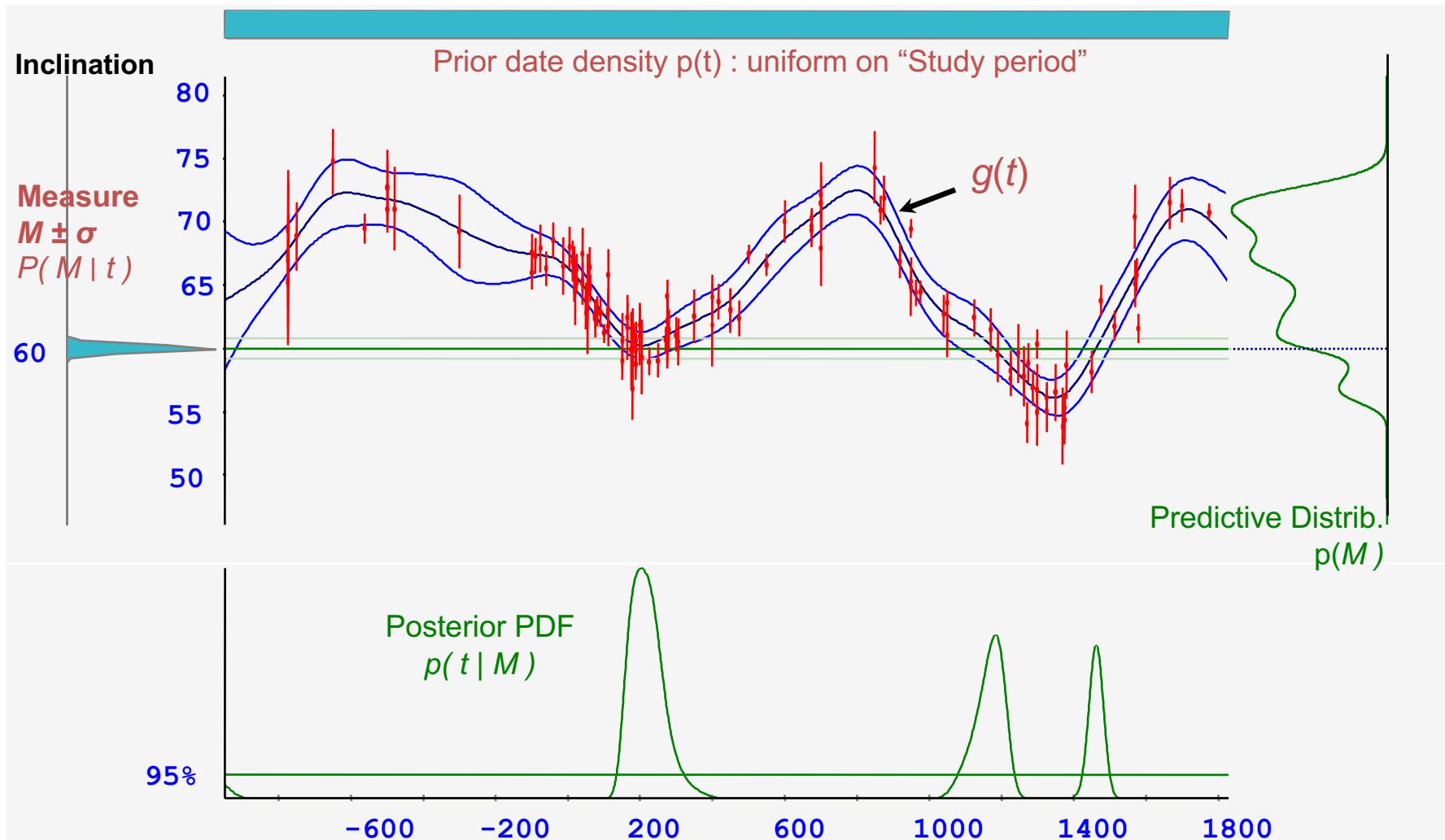
Chronometric dating : measurement / age determination

Radiocarbon calibration



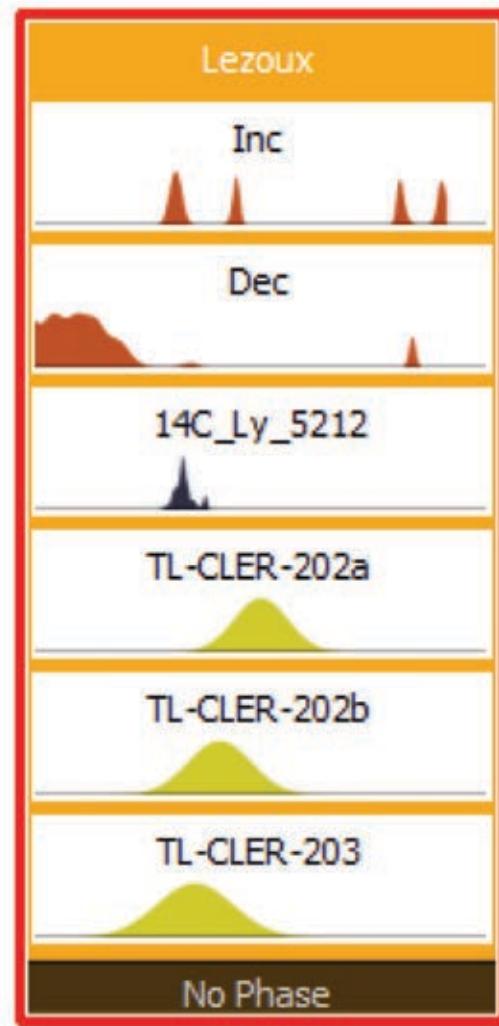
Archaeomagnetic calibration

Measurement Calibration → calendar Date



Event date model in *ChronoModel*

- *ChronoModel* est basé sur le concept de date d'événement θ (Event date).
- Un Fait est un événement « ponctuel » dans le temps pour lequel on peut définir un modèle statistique de type bayésien hiérarchique.
- Ce modèle est « robuste » car il fait intervenir des variances individuelles (erreurs humaines, d'échantillonnage, facteurs inconnus) qui gèrent les outliers.
- Il est estimé grâce à des datations effectuées en laboratoire ou issues de références (par exemple typo-chronologiques) supposées contemporaines de l'événement à dater.



Event model and associated DAG

- Likelihood :

$$p(M_1, \dots, M_n | t_1, \dots, t_n) = \prod_{i=1}^n \frac{1}{s_i \sqrt{2\pi}} e^{\frac{-1}{2s_i^2} (M_i - g_i(t_i))^2}$$

- Prior on the dates t_i

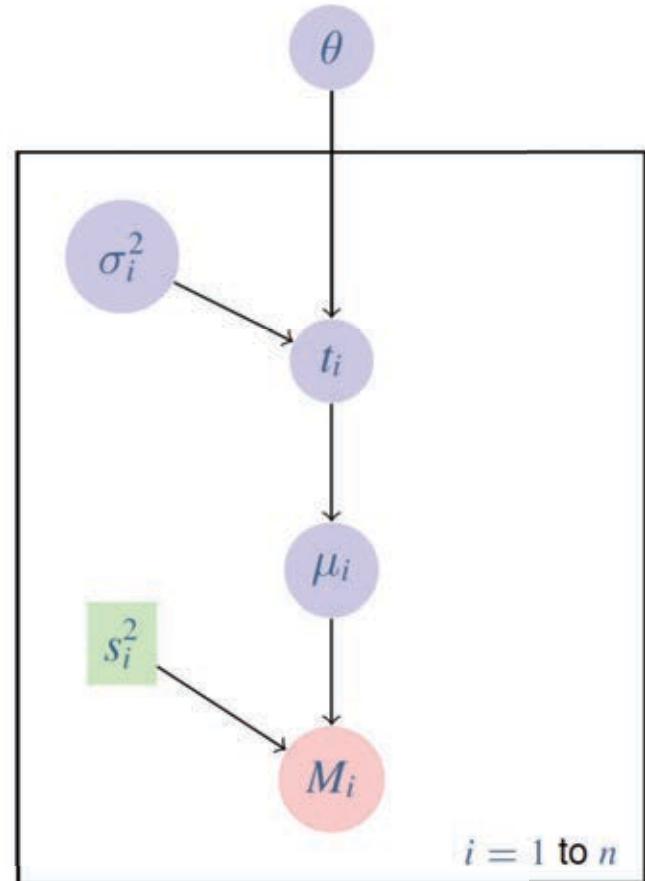
$$p(t_1, \dots, t_n | \theta, \sigma_1^2, \dots, \sigma_n^2) = \prod_{i=1}^n \frac{1}{\sigma_i \sqrt{2\pi}} e^{\frac{-1}{2\sigma_i^2} (t_i - \theta)^2}$$

- Prior on the parameter of interest θ :

$$p(\theta) \propto 1_T(\theta) \text{ where } T = [T_a, T_b]$$

- Prior on the individual variances :

$p(\sigma_1^2, \dots, \sigma_n^2)$: Shrinkage distribution

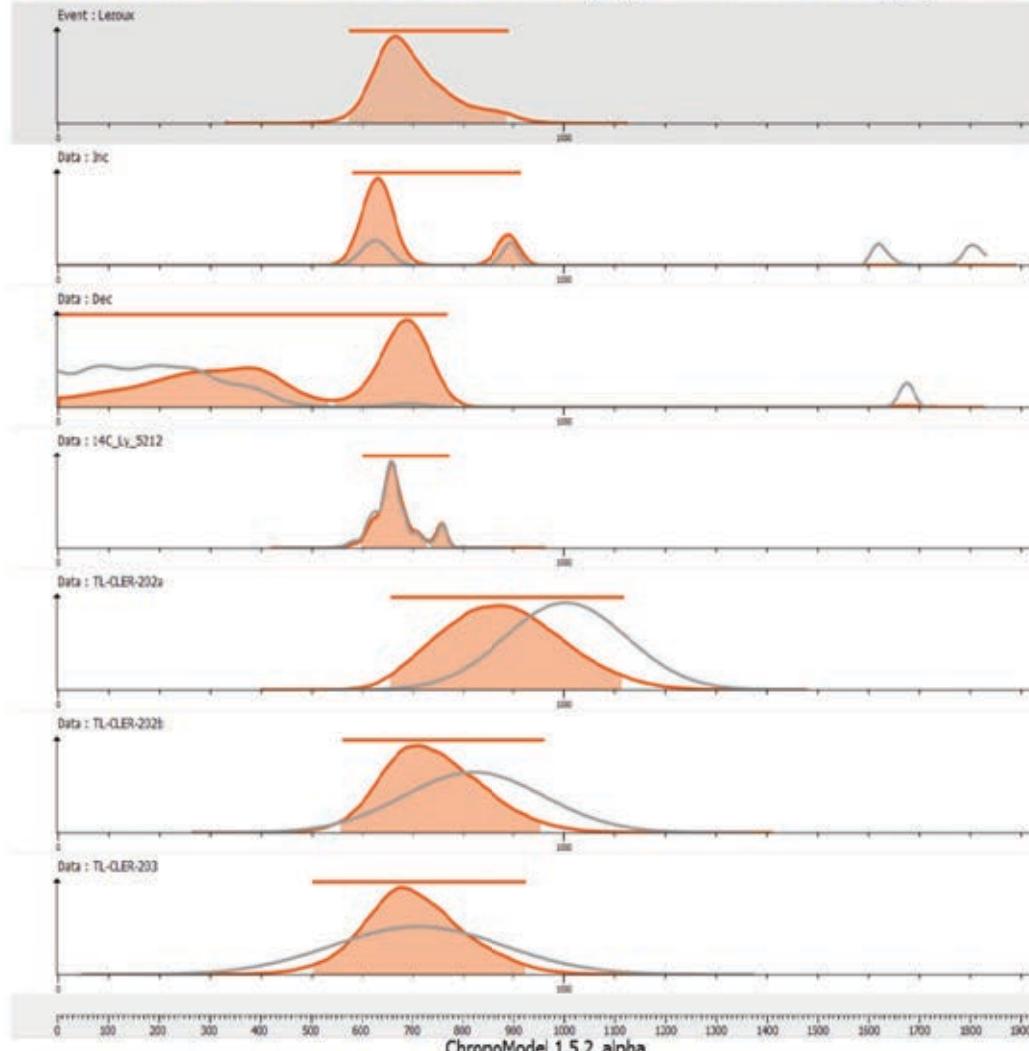


This distribution is rather diffuse in the sense that its variance and mean are infinite
So it can be considered weakly informative.

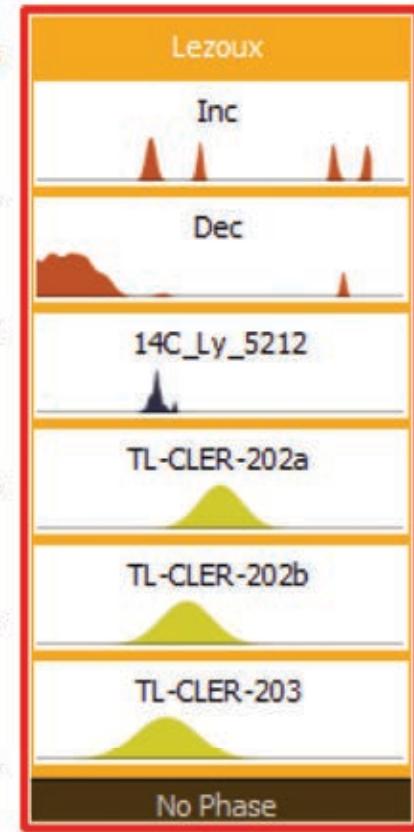
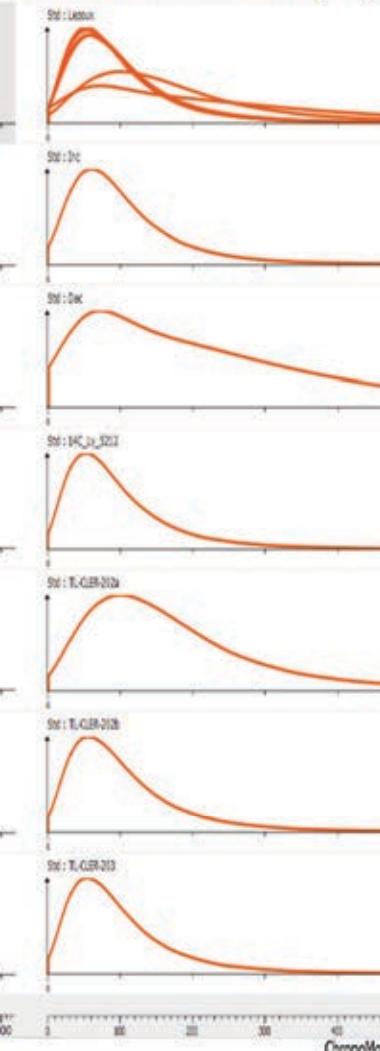
An event example: last firing of a kiln

Posterior event date and indiv. errors distributions

Posterior densities : Event (θ) and dates (t_i)



Individual std (σ_i)

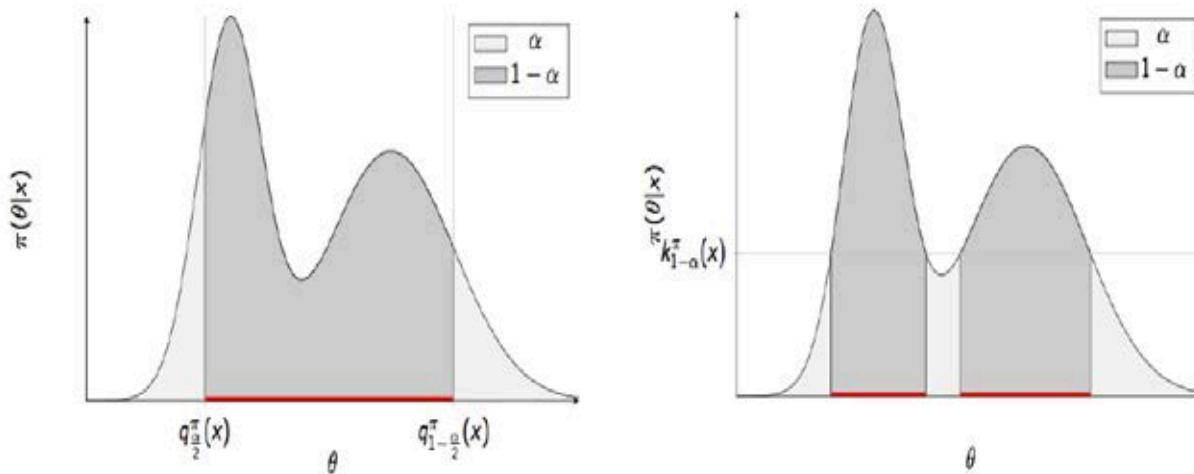


How to summarize statistical posterior results ?

Since Buck's works in the 90s,

The distribution is summarized as a « **posterior probability density** » region **HPD** at a given confidence level: generally $1-\alpha = 95\%$

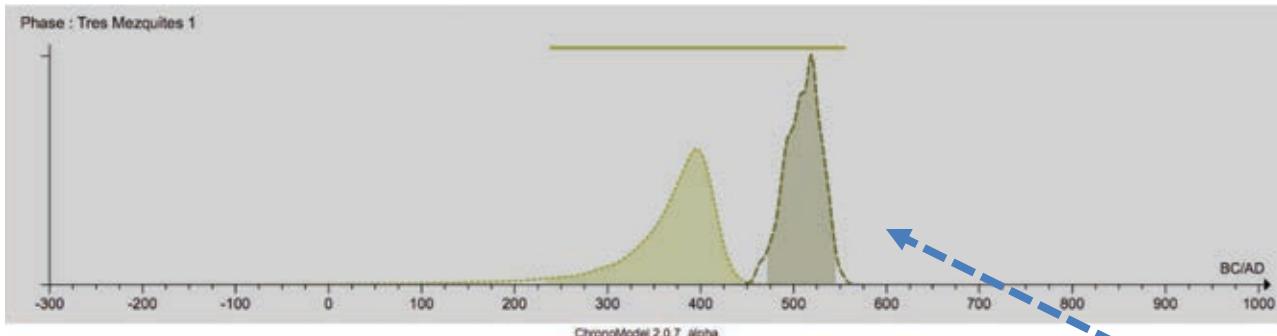
The **credibility interval** (CI) may also be used.



Moreover, it is possible to use **pointwise estimates: MAP** (the **Mean** is not advisable)

Phases = group of Events

Start, End, Phase range, Duration estimates



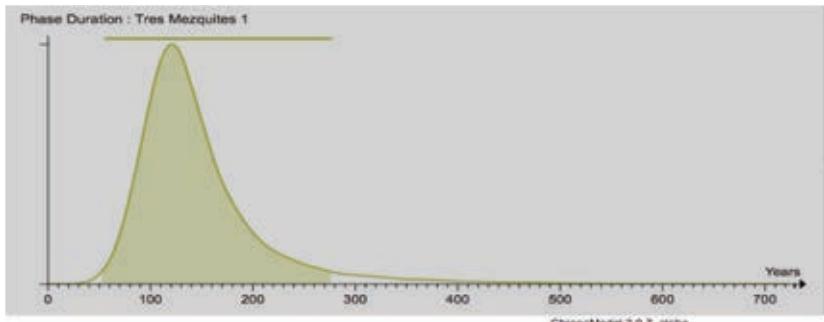
Begin (posterior distrib.)

MAP = 395

HPD Region (95 %) : [233 ; 442] (95%) BC/AD

Credibility Interval (95 %) : [237 ; 441] BC/AD

Phase Time Range (95 %) : [239 ; 553] BC/AD



Duration (posterior distrib.)

MAP = 121

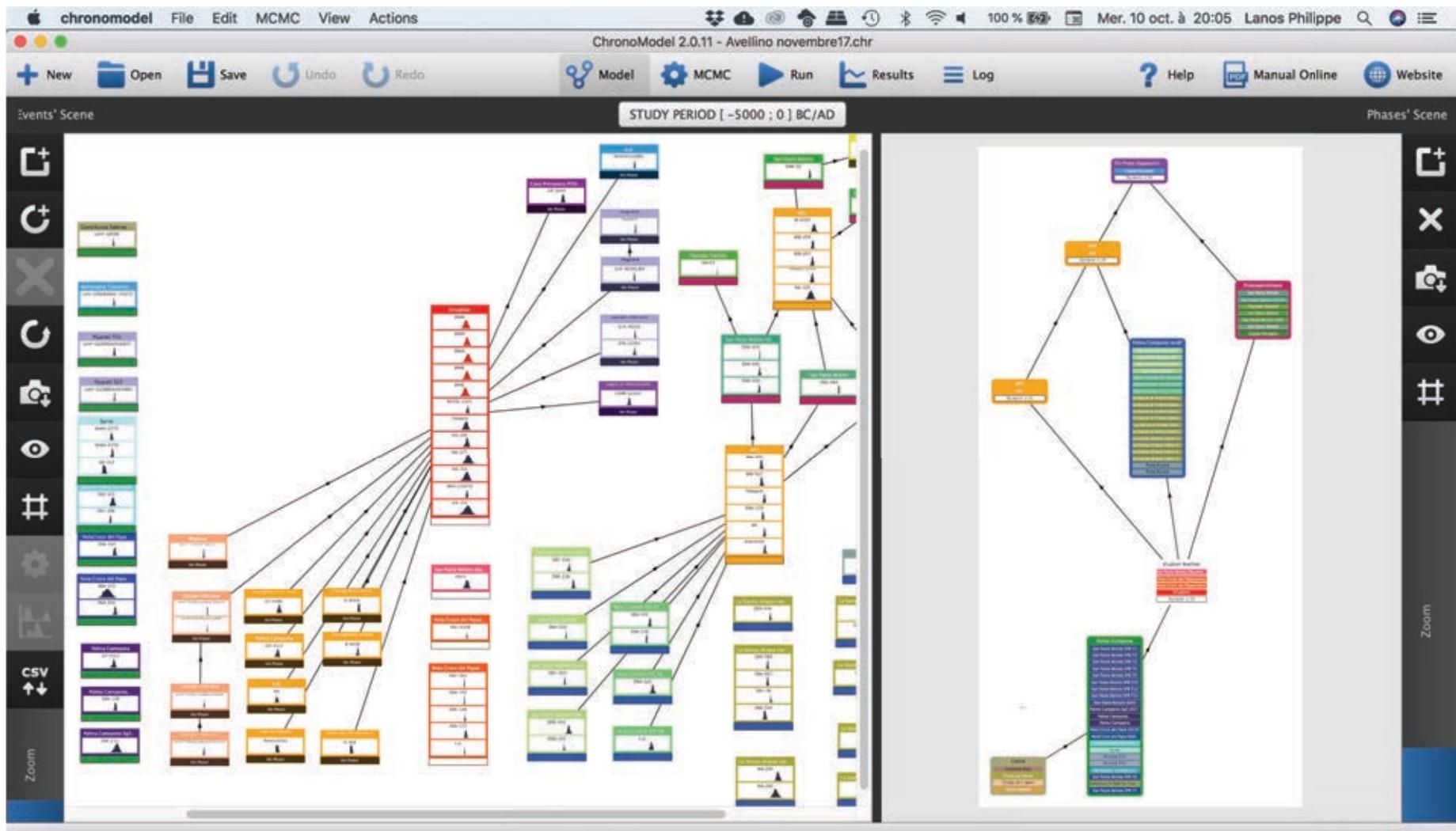
HPD Region (95 %) : [53 ; 275] (95%)

Credibility Interval [57 ; 275] AD

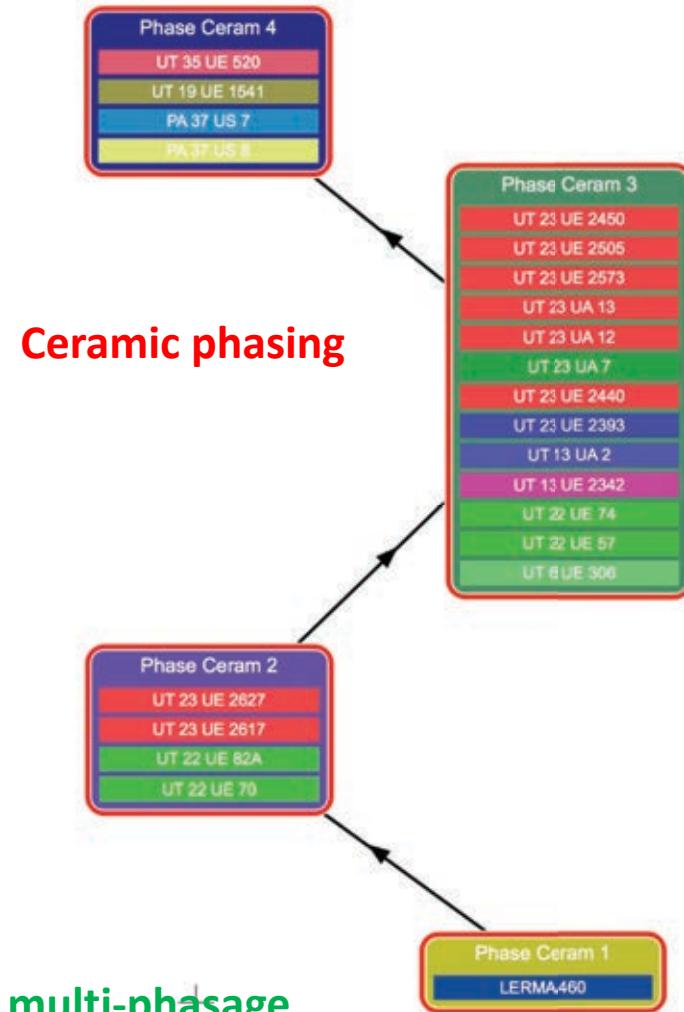
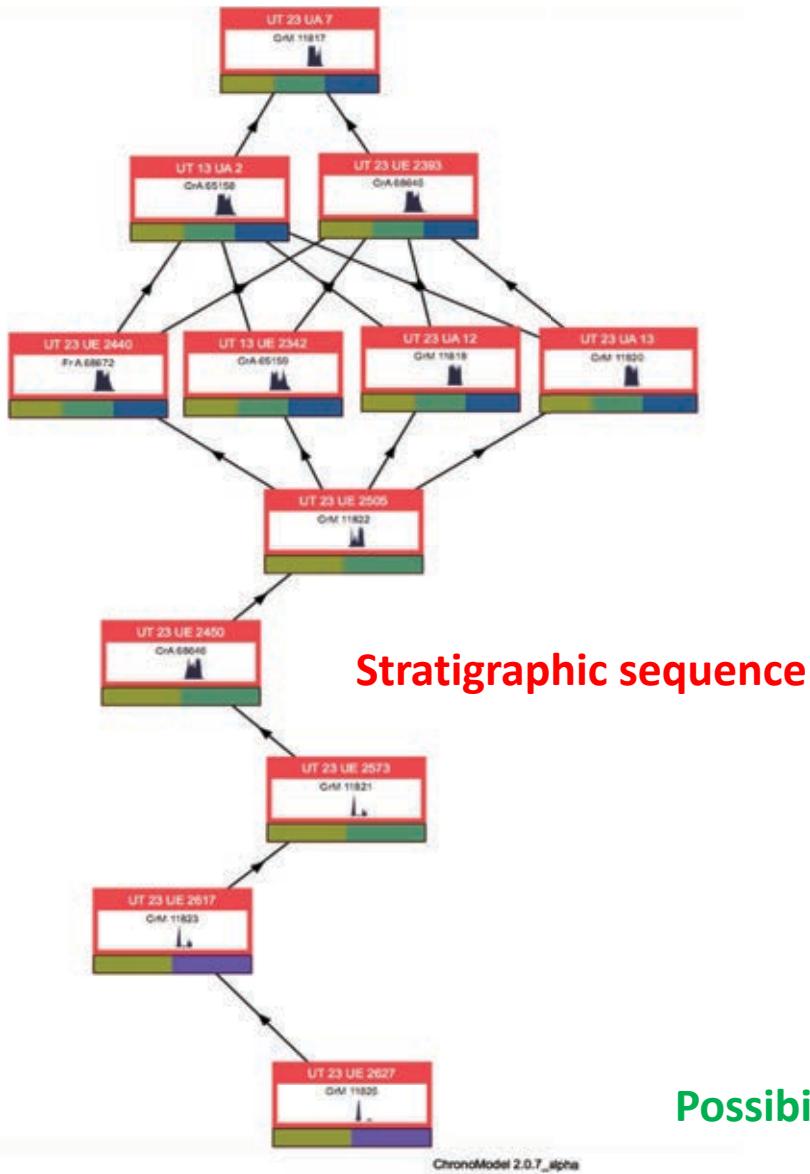
Tres Mezquites 1
UT 23 UE 2627
UT 23 UE 2617
UT 23 UE 2450
UT 23 UE 2505
UT 23 UE 2573
UT 23 UA 13
UT 23 UA 12
UT 23 UA 7
UT 23 UE 2440
UT 23 UE 2393
UT 13 UA 2
UT 13 UE 2342

Events et Phases dans ChronoModel

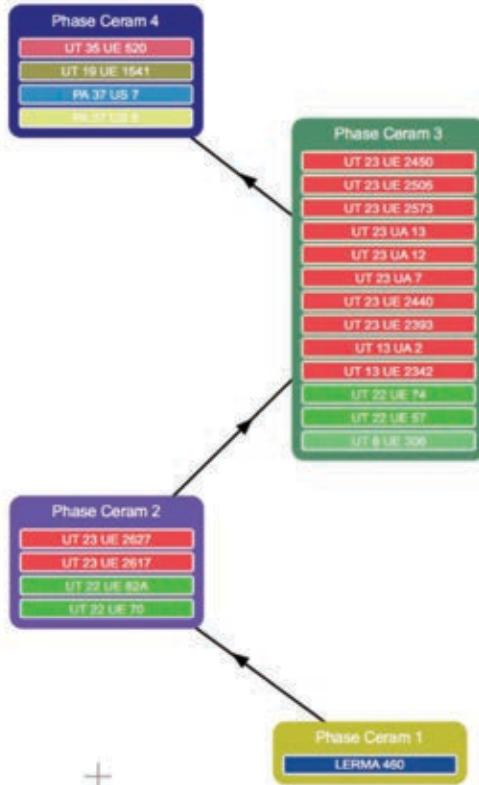
- Fast and intuitive model creation.
- Import/export CSV data files with drag & drop.
- 3 levels models : Phase / Event / Data.
- Long study periods support.



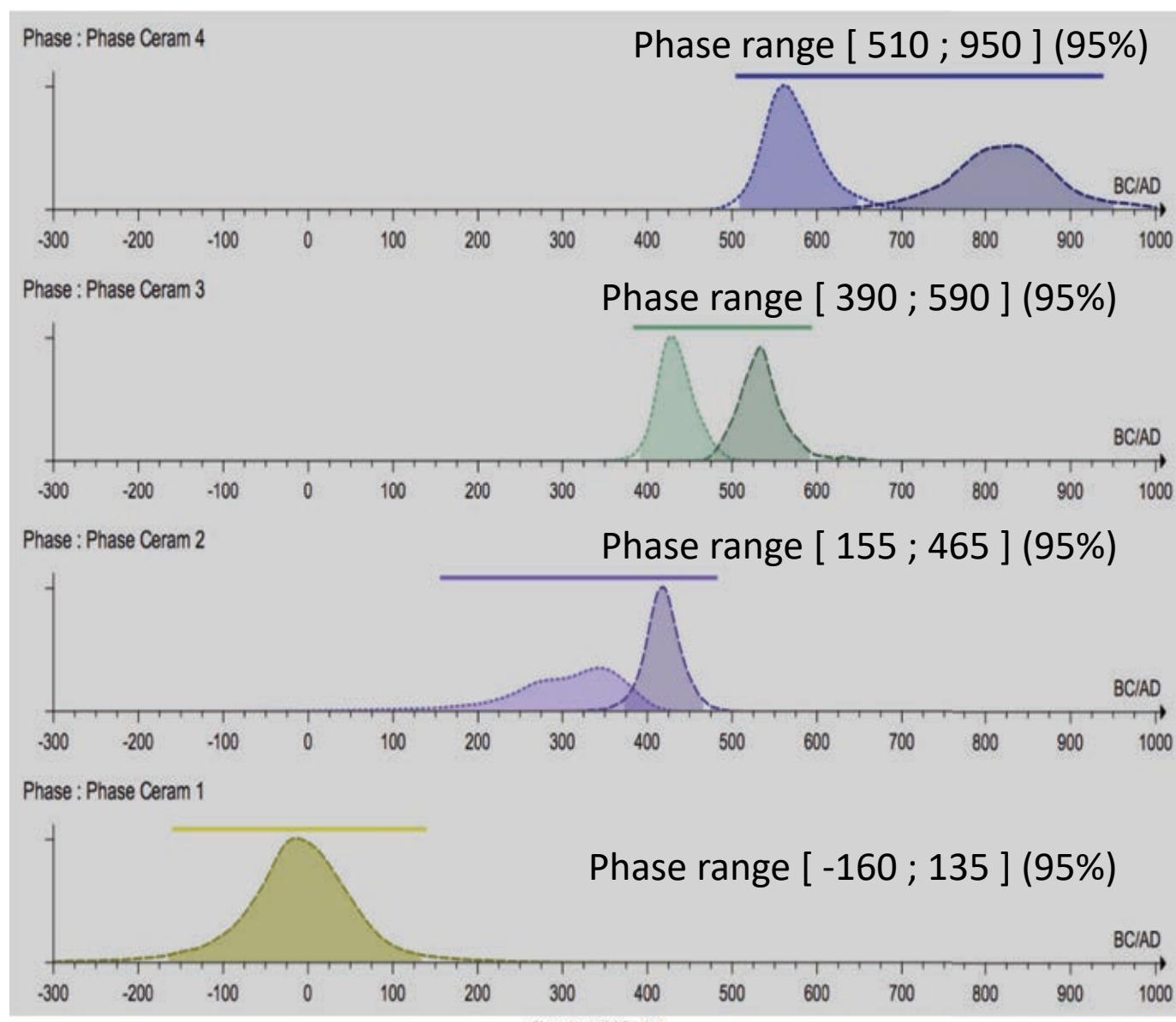
Stratigraphic / succession constraints



Chronology of the Ceramic phases



Ceramic seriation
Succession constraint
Transition overlap :
proof of continuity



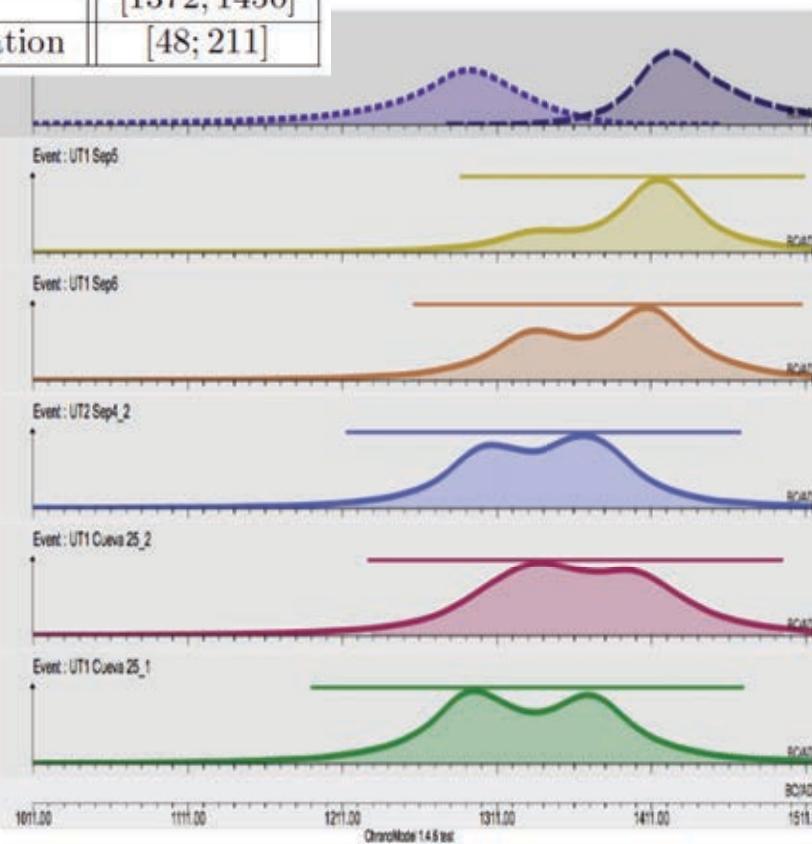
Prior information on phase duration

Site of Malpais Prieto (Michoacan State, Mexico) (Pereira et al., 2016)

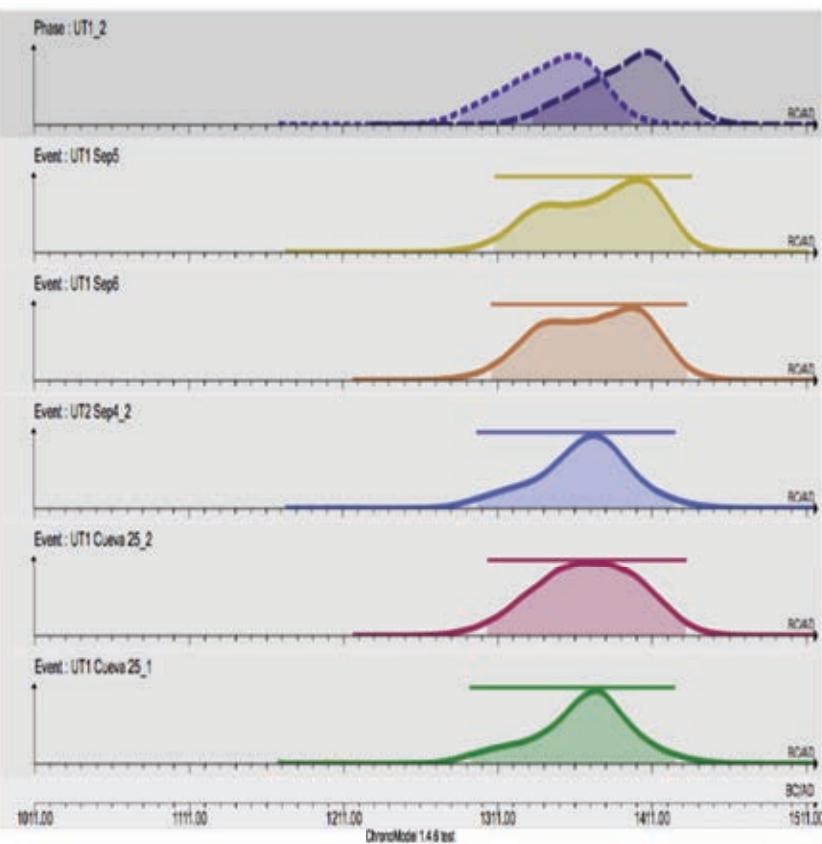
Occupation span ≤ 60 ans

Comparison of HPD regions and posterior densities

PHASE	
Begin	[1209; 1355]
End	[1372; 1450]
Duration	[48; 211]



PHASE ($\tau_0 = 60$)	
Begin	[1286; 1397]
End	[1335; 1443]
Duration	[28, 60]

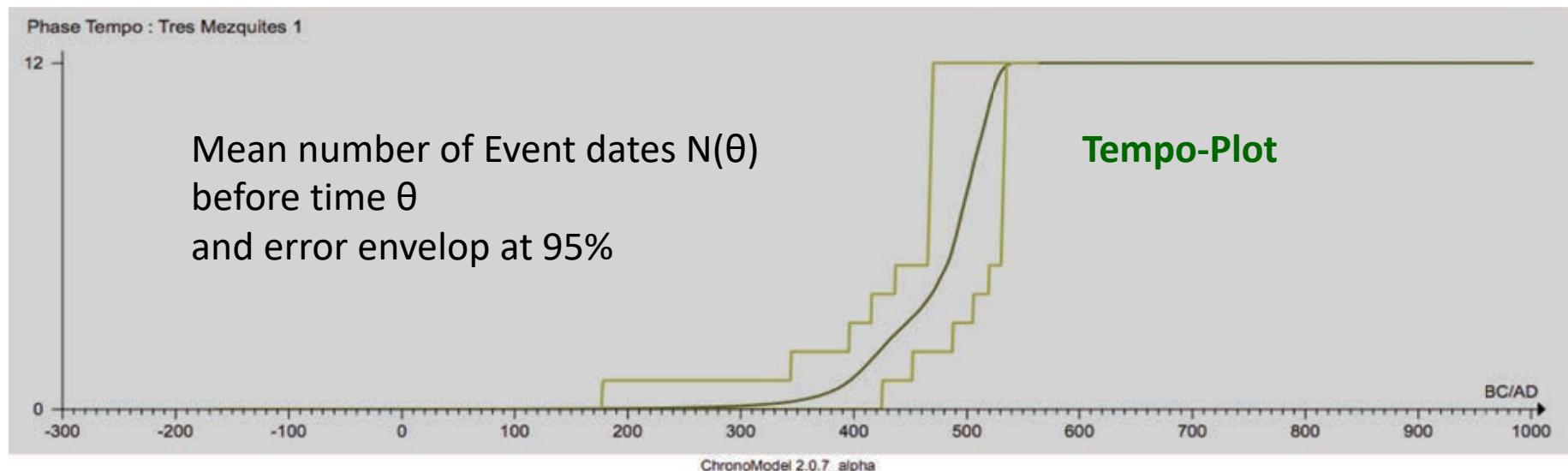


How to Characterize rythms : Tempo-Plot and Activity curves

On peut assimiler l'occurrence des événements (à partir de l'estimation des dates d'événement) à un processus de comptage en traitement du signal.

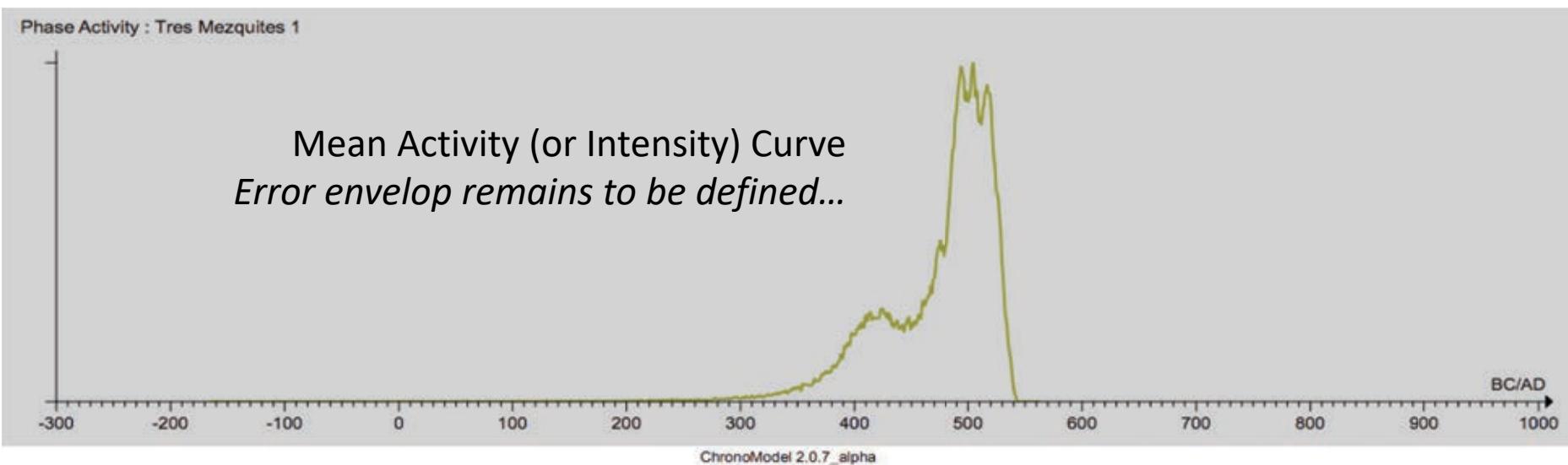
Le but est d'estimer la fonction d'intensité $\lambda(\theta)$ qui mesure la probabilité instantanée d'un événement.

Le **Tempo-Plot** donne le nombre moyen d'événement $N(\theta)$ qui se sont produits avant une date θ . On peut définir une enveloppe de crédibilité autour de cette courbe.



Activity (or intensity) Curve

La dérivée (taux de variation) du Tempo-Plot donne la fonction d'intensité $\lambda(\theta)$, c'est-à-dire le nombre d'événement par unité de temps. On montre que cette intensité est aussi donnée par la somme des densités de date a posteriori des événements considérés. Mais s'agissant de données bruitées, il faut aussi définir une enveloppe d'erreur.



Toute la question reste de savoir si les événements considérés sont représentatifs de tous les événements qui caractérisent le phénomène étudié.
Attention aux sur-interprétations !!!



Merci de votre attention

Formations ChronoModel CM 2.0

- Formation CNRS ANF, SBR Roscoff, novembre 2018
- Formation interne INRAP Hauts-de-France, novembre 2018

Bibliographie

Lanos Ph. and Philippe A., 2017. Hierarchical Bayesian modeling for combining dates in archaeological context. *Journal de la Société Française de Statistique*, Vol. 158 (2) p. 72-88. <http://journal-sfds.fr/index.php/J-SFdS/issue/view/68>

Lanos Ph. and Philippe A., 2018. Event date model: a robust Bayesian tool for chronology building. *Communications for Statistical Applications and Methods (CSAM)*, Vol. 25, No. 2, 131–157. <https://doi.org/10.29220/CSAM.2018.25.2.131>