



HAL
open science

**gestion d'imperfection de données temporelles :
construction de distributions chronologiques par pas de
temps, à partir d'événements de datation imprécise ou
de précision hétérogène**

Bruno Desachy

► **To cite this version:**

Bruno Desachy. gestion d'imperfection de données temporelles : construction de distributions chronologiques par pas de temps, à partir d'événements de datation imprécise ou de précision hétérogène. Atelier "des données aux plateformes géohistoriques en accès libre", Jun 2021, Abbaye de Royaumont, (Asnières-sur-Oise, Val d'Oise), France. hal-04090887

HAL Id: hal-04090887

<https://hal.science/hal-04090887v1>

Submitted on 9 May 2023

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.

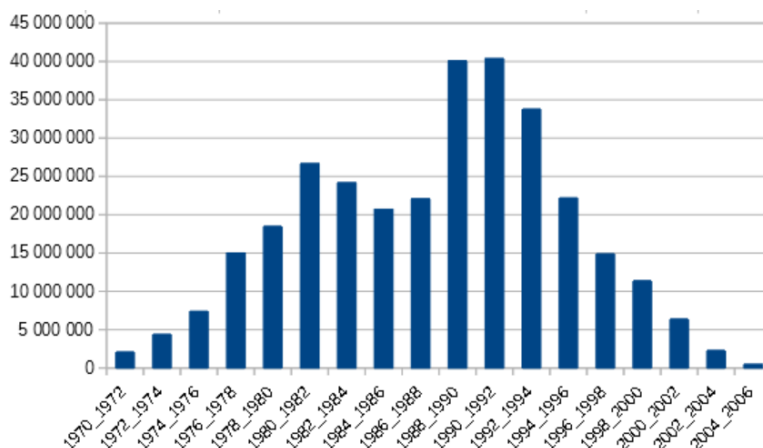
L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Mardi 29 juin 14h - 14h45 - Bruno Desachy :

gestion d'imperfection de données temporelles : construction de distributions chronologiques par pas de temps, à partir d'événements de datation imprécise ou de précision hétérogène

présentation :

Pour une vision synthétique des évolutions techniques, fonctionnelles, économiques, sociales, etc., les archéologues sont amenés, comme les historiens et les économistes, à construire des courbes de comptages de différents événements, par pas de temps (tous les siècles, demi-siècles, quarts de siècles, décennies, etc.), en choisissant ce pas de temps, correspondant à la résolution chronologique souhaitée. La figure ci-dessous est par exemple la courbe d'évolution (distribution chronologique par pas de temps) des ventes de musique enregistrée sur cassette audio de 1970 à 2006, avec un pas de temps bisannuel¹



Cet exemple contemporain ne pose pas de problèmes d'imperfection de données : les ventes sont connues à l'année, les comptages pour obtenir cette distribution avec ce pas de temps de deux ans sont donc sans difficultés.

Dans d'autres cas, les sources anciennes (ou récentes...) sont moins simples à traiter car imprécises (intervalle d'imprécision – d'un texte mal daté par exemple – excédant le pas de temps envisagé ou tombant à cheval entre deux pas de temps), et surtout différentiellement imprécises (par exemple entre une datation dendrochronologique dans les meilleures conditions c'est à dire au semestre près, et une datation radiocarbone tombant dans un palier de la courbe de calibration, avec deux bons siècles d'imprécision...).

Comment tenir compte de ces imprécisions pour construire ce type de distributions chronologiques avec les données imparfaites dont nous devons souvent nous contenter ?

Les solutions « violentes » consistant ne conserver que les données les plus précises, ou à « noyer » toutes les données dans des pas de temps correspondant aux intervalles d'imprécision les plus larges, ne sont pas satisfaisantes, car la première conduit à fortement restreindre les corpus, et la seconde à beaucoup perdre en résolution chronologique.

1 source : Département Etudes Prospectives Statistiques, ministère de la Culture :
http://www.culture.gouv.fr/Thematiques/Etudes-et-statistiques/Statistiques-culturelles/Donnees-statistiques-par-domaine_Cultural-statistics/Musique-enregistree

Des solutions plus nuancées existent : lorsque l'intervalle d'imprécision de datation d'un événement excède le pas de temps choisi, l'imprécision se transforme en incertitude (d'occurrence ou d'existence de cet événement) dans les pas de temps concerné, incertitude que l'on peut quantifier de façon probabiliste ou possibiliste. La distribution obtenue est alors formée de cumuls de probabilités ou de possibilités et non plus de simples sommes d'unités (correspondant aux événements concernés) par pas de temps ; mais on peut pareillement interpréter cette distribution en termes d'évolution du phénomène étudié. Les calculs peuvent cependant être fastidieux si il y a un grand nombre d'événements à traiter.

Ce problème a été discuté dans le cadre de l'atelier SITraDA² car il concernait plusieurs thèses alors en cours. Il nous a mené à la réalisation d'un outil simple (application « Chronophage », constituée de macro-commandes exécutables dans le tableur de la suite logicielle libre LibreOffice³) permettant de construire de telles distributions à partir d'événements frappés d'imprécisions différentielles, suivant deux cas de figure principaux :

- distribution chronologique par pas de temps à partir d'événements réduits à des instants (ou munis d'une durée considéré comme négligeable à l'échelle de temps choisie) ; par exemple naissances ou apparitions, morts ou disparitions, émissions d'actes ou documents historiques, dépôts archéologiques formant des « ensembles clos » (dépôts primaires constitués dans un temps très bref).
- distribution chronologique par pas de temps à partir d'événements munis d'une durée (connue ou estimée) ; par exemple durées d'occupation de sites, de production ou circulation de types de traits ou d'objets matériel...

Deux approches simples pour gérer l'imprécision dans ces cas de figure seront présentées et mises en oeuvre avec l'application Chronophage : une répartition de nature probabiliste de l'imprécision, et une approche subjective par valeurs de vérité directement attribuées par le chercheur.

L'exploitation possible des distributions chronologiques obtenues sera aussi évoquée : représentations graphiques (courbe, histogrammes, sériographes), traitement multidimensionnel d'un ensemble de distributions chronologiques.

Les personnes souhaitant participer à l'atelier en effectuant les manipulations pratiques devront être munies d'un ordinateur et du logiciel LibreOffice (téléchargeable gratuitement : <https://fr.libreoffice.org/>). les travaux auront lieu avec des jeux de données de démonstrations mais aussi avec les données des participants si ceux-ci souhaitent travailler avec leurs propres données.

2 Atelier Systèmes d'Information et TRAitement de Données Archéologiques ; séminaire doctoral de l'ED 112 (Paris 1) et du programme ArchéoFab de l'UMR 7041ArScAn - <https://sitrada.hypotheses.org/>

3 <https://abp.hypotheses.org/4284>