

Sujet de thèse

Titre de la thèse :

Modélisation bayésienne sous contraintes de monotonie appliquée à des données de géochronologie à haute résolution.

Bayesian modelling under monotonicity constraints applied to high-resolution geochronology data.

Mots clés : Modélisation ; Statistique Bayésienne ; Application en datation (Archéologie) ; Régression sous contraintes ; information a priori ; données à haute résolution.

Key words: Modelling; Bayesian statistics; application to dating (archaeology); constrained regression; prior information; data high-resolution.

Directrice de thèse : Anne Philippe (PR NU, LMJL)

Co encadrant : Guillaume Guérin (CR CNRS, Géosciences Rennes)

Début de la thèse : Septembre 2024

Ces dernières années, la modélisation bayésienne pour l'analyse des données de datation en archéologie et en sciences de la terre a connu un essor important. Elle fournit des outils statistiques pour prendre en compte les mesures physiques, leurs incertitudes, et les informations a priori venant du contexte archéologique. La construction de la loi a priori joue un rôle central, elle permet en particulier de prendre en compte les contraintes d'ordre entre les paramètres du modèle (voir [1] par exemple). Par exemple, la connaissance de la stratigraphie implique des contraintes d'ordre sur les dates à estimer.

Dans cette situation, la modélisation chronologique bayésienne naturelle proposée dans [2], [3] [4] et [5] consiste à choisir comme loi a priori pour n dates à estimer $(\theta_1, \dots, \theta_n)$, la loi d'une statistique d'ordre uniforme c'est à dire la loi uniforme sur le simplexe de $[0,1]^n : \{0 < \theta_1 < \dots < \theta_n < 1\}$.

Comme illustré par exemple dans [6], pour ce choix de loi a priori, les estimateurs de Bayes ne sont pas consistants lorsque le nombre de dates n tend vers l'infini ou lorsque que des dates sont égales ou contemporaines. Ce modèle est donc mis en défaut pour un échantillonnage dit à haute

résolution, au sens où la différence entre deux événements datés est petite par rapport aux incertitudes de mesures.

La thèse proposée consistera à proposer des modèles alternatifs pour corriger ce problème dans le contexte des données « haute résolution ».

La première partie sera consacrée à la construction de nouvelles lois a priori et à la validation asymptotique des estimateurs de Bayes associés. Cette partie s'appuiera sur le travail exploratoire que nous avons réalisé dans [7]. L'approche repose sur les propriétés des statistiques d'ordre que l'on perturbe pour gagner en robustesse et en précision dans le contexte des données « haute résolution ».

Dans la seconde partie, l'idée principale est d'intégrer dans la modélisation chronologique l'information complémentaire que nous possédons sur la profondeur de chaque échantillon daté. Plus généralement, ce problème consiste à construire une loi a priori sur le paramètre θ dépendant d'une variable exogène E sous l'hypothèse que $\theta = g(E)$ où g est inconnue mais supposée monotone. L'objectif de cette partie ne sera pas d'estimer la fonction g mais d'améliorer l'estimation de θ en termes de réduction de biais et de précision en apportant de l'information via la variable E .

La troisième partie portera sur l'estimation de la relation âge-profondeur dans une stratigraphie, c'est à dire de la fonction g à partir de la loi a posteriori de θ . L'objectif est de construire une prévision en loi de l'âge d'une couche géologique non datée. Ce problème nécessitera d'estimer la fonction g sous contrainte de monotonie. Une piste pour aborder ce problème est la régression bayésienne par processus gaussien ; elle permet d'intégrer des contraintes physiques ou d'autres informations a priori dans la régression (voir [8]).

Les résultats obtenus dans les parties précédentes seront mis en œuvre sur des données de datation par OSL en lien avec le projet ERC Quina World (porteur G. Guérin). Le site de Jonzac (Charente-Maritime) est un site en pied de falaise présentant une des stratigraphies les plus importantes de la région pour le Paléolithique moyen. La modélisation bayésienne « haute fréquence » permettra alors d'apporter un éclairage nouveau sur la durée d'occupation cumulée du site par les groupes de Néandertaliens, mais aussi et surtout sur la cyclicité des passages de groupes humains pour des activités cynégétiques. Plus généralement, les modèles développés pourront servir à toute étude nécessitant une grande résolution chronologique, comme les recherches sur les fluctuations climatiques passées.

Références :

- [1] Gelfand, A E., A. F. M. Smith, et T.-M. Lee. «Bayesian Analysis of Constrained Parameter and Truncated Data Problems Using Gibbs Sampling.» *Journal of the American Statistical Association*, 1992: 87:418, 523-532.
- [2] Lanos, P., and A. Philippe. 2018. “Event Date Model: A Robust Bayesian Tool for Chronology Building.” *Communications for Statistical Applications and Methods* 25 (2): 131–57.
- [3] Ramsey, C. B. «Bayesian analysis of radiocarbon dates. .» *Radiocarbon*, 2009: 51(1), 337-360.
- [4] Combès, Benoit, and Anne Philippe. 2017. “Bayesian Analysis of Individual and Systematic Multiplicative Errors for Estimating Ages with Stratigraphic Constraints in Optically Stimulated Luminescence Dating.” *Quaternary Geochronology* 39: 24–34.
- [5] Philippe, Anne, Guillaume Guérin, and Sebastian Kreutzer. 2019. “BayLum - an R Package for Bayesian Analysis of Osl Ages: An Introduction.” *Quaternary Geochronology* 49: 16-24,
- [6] Steier, P., & Rom, W. (2000). The use of Bayesian statistics for 14C dates of chronologically ordered samples: a critical analysis. *Radiocarbon*, 42(2), 183-198.
- [7] Guillaume Guérin ; Pierre Guitton-Boussion; Imène Bouafia; Anne Philippe The issue of modelling stratigraphic constraints in a Bayesian framework: the case study of the East Mound of Çatalhöyük (Turkey) *Soumis* 2023
- [8] Swiler, L., M. Gulian, A. Frankel, C. Safta, et J. Jakeman. «A Survey of Constrained Gaussian Process: Approaches and Implementation Challenges.» *Journal of Machine Learning for Modeling and Computing*, 2020: 1(2):119–156.