

Proposition de sujet de stage

Modélisation par chaînes de Markov couplées de la dynamique conjointe d'occurrences de séismes

Contexte. Le stage concerne la modélisation spatio-temporelle de l'occurrence des séismes, dans le cadre des Chaînes de Markov Cachées (Hidden Markov Models, HMM en anglais) pour la dynamique temporelle, couplées entre des lieux différents d'une même province géologique pour l'aspect spatial. Il s'agira de formaliser un ou plusieurs modèles, inférer des paramètres d'intérêt sur des données réelles, et interpréter les résultats. Des modélisation dans le cadre des HMM de la dynamique d'occurrence des séismes en un lieu géographique ont déjà été considérées dans les articles suivants: [2] et [3]. Les observations sont les occurrences des séismes avec leur magnitude. La chaîne cachée permet de modéliser les différents régimes par lequel le niveau de stress sismique dans la lithosphère passe au cours du temps, et qui ne sont pas observés. Lorsque le stress est trop élevé la probabilité d'occurrence d'un séisme est forte, ce qui conduit à un relâchement du stress et donc à un changement de régime. Cependant ces modèles ne tiennent pas compte de l'influence du niveau de stress dans les régions voisines, sur le niveau de stress local. C'est ce que nous proposons de modéliser dans le cadre de ce stage, en couplant les dynamiques des HMM de chaque site dans le cadre des HMM couplées (CHMM, [1]).

Description du projet. Le déroulé du stage comprend trois grandes étapes : modélisation, exploration et sélection d'algorithmes pour l'inférence du modèle et application aux données réelles.

1. **Modélisation.** Après une étude bibliographique des travaux existants sur la modélisation de la dynamique de séismes par HMM, il s'agira d'une part de faire un choix sur la nature des variables cachées et des observations. Utiliser directement les données brutes n'est pas possible car sur un pas de temps plusieurs séismes peuvent avoir lieu. Il s'agira de proposer comment résumer cette information en une ou plusieurs variables pour lesquelles il sera simple de modéliser la distribution conditionnellement au régime. Il s'agira ensuite de construire un modèle CHMM intégrant le couplage entre les deux lieux (chaînes couplées) ainsi que les boucles de rétro-action entre séismes observés et changement d'état du stress. En particulier il faudra proposer une modélisation paramétrique pour la probabilité de transition du régime caché d'une chaîne sachant celui de l'autre chaîne.
2. **Inférence.** Pour l'inférence des paramètres du modèle CHMM et des régimes cachés, les approches bayésiennes seront privilégiées car elles fournissent non seulement les valeurs des paramètres, mais une distribution a posteriori. L'étudiant(e) explorera plusieurs options (Metropolis-Hasting, Gibbs Sampling, méthodes Hamiltoniennes, algorithmes de Langevin sur Variétés Riemanniennes). Des algorithmes ont déjà été proposés pour des CHMM appliqués la modélisation de maladies infectieuses et pourront être un point de départ. Il

s'agira de mettre en oeuvre la méthode qui semblera la plus adaptée à la complexité du modèle. Le comportement de l'algorithme d'inférence sera testé sur des données simulées. On pourra également tester si le modèle CHMM améliore la qualité de prédiction des observations et des régimes latents ainsi que l'inférence des paramètres, par rapport à un HMM.

3. **Application.** Le modèle CHMM sera appliqué aux données réelles d'occurrence de séismes en deux localisations différentes, en Grèce. Cela permettra par exemple de quantifier l'importance du couplage tant dans l'évolution du stress local que dans l'occurrence des séismes.

Profil. Etudiant(e) préparant un Master 2 ou en dernière année d'école d'ingénieur avec a) une spécialisation en statistique, probabilités, science des données ; b) des bases solides en programmation sous R ; c) un intérêt pour les applications en science de la vie et de la terre.

Encadrants. Le stage sera encadré par Sébastien Coube et Nathalie Peyrard (MIAT - INRAE Toulouse), Irène Votsi (LIEC, Université de Lorraine) et Alain Franc (BioGeCo, INRAE Bordeaux).

Lieu et durée. Le stage se déroulera au sein de l'unité de Mathématiques et Informatiques du centre Occitanie-Toulouse d'INRAE (MIAT, <https://miat.inrae.fr/>). Le stage peut débuter à partir du 1er avril et pour une durée de 5 à 6 mois.

Contact. Merci d'envoyer votre candidature, avec un CV à sebastien.coube@inrae.fr, nathalie.peyrard@inrae.fr, Eirini.Votsi@univ-lorraine.fr et alain.franc@inrae.fr

References

- [1] M. Brand. Coupled hidden Markov models for modeling interacting processes. Technical Report 405, MIT Media Laboratory, Cambridge, 1997.
- [2] K. Orfanogiannaki, D. Karlis, and G. A. Papadopoulos. Identifying seismicity levels via poisson hidden markov models. *Pure and Applied Geophysics*, 167(8):919–931, 2010.
- [3] I. Votsi, N. Limnios, G. Tsaklidis, and E. Papadimitriou. Hidden markov models revealing the stress field underlying the earthquake generation. *Physica A: Statistical Mechanics and its Applications*, 392(13):2868–2885, 2013.