

# Inférence statistique dans les modèles espace-états

Thị Tuyết Trang Châu, Pierre Ailliot, Valérie Monbet, Pierre Tandeo, Juan Jose Ruiz

Les méthodes d'assimilation de données séquentielles sont généralement formalisées sous la forme d'un modèle espace-état, dans lequel on distingue le modèle dynamique qui décrit l'évolution du processus physique (état), et le modèle d'observation qui décrit le lien entre le processus physique et les observations disponibles.

Trois problèmes liés à l'inférence statistique pour les modèles espace-états seront abordés dans cet exposé.

- **Reconstruction de l'état.** Nous introduirons tout d'abord un algorithme de lissage original qui combine les algorithmes Conditional Particle Filter (Andrieu et al., 2010) et Backward Sampling (Lindsten et al., 2013). Cet algorithme CPF-BS permet une exploration efficace de l'état de la variable physique, en raffinant séquentiellement l'exploration autour des trajectoires qui respectent le mieux les contraintes du modèle dynamique et des observations. Nous montrerons sur un modèle jouet (Lorenz 63) que, à temps de calcul égal, l'algorithme CPF-BS donne de meilleurs résultats que l'algorithme EnKS stochastique qui est couramment utilisé dans les applications opérationnelles (cf Chau et al., 2018) .
- **Estimation des paramètres.** Nous aborderons ensuite le problème de l'estimation des paramètres inconnus dans les modèles espace-état. L'algorithme le plus usuel en statistique pour estimer les paramètres d'un modèle espace-état est l'algorithme EM qui permet de calculer itérativement une approximation numérique des estimateurs du maximum de vraisemblance. Nous montrerons que les algorithmes EM et CPF-BS peuvent être combinés efficacement pour estimer les paramètres d'un modèle jouet.
- **Remplacement du modèle dynamique par un émulateur construit à partir de données.** Pour certaines applications, le modèle dynamique est inconnu ou très coûteux à résoudre numériquement mais des observations ou des simulations sont disponibles. Il est alors possible de reconstruire l'état conditionnellement aux

observations en utilisant des algorithmes de filtrage/lissage dans lesquels le modèle dynamique est remplacé par un émulateur statistique construit à partir des observations (Lguensat et al., 2017.). Nous montrerons que les algorithmes EM et CPF-BS peuvent être adaptés dans ce cadre et permettent d'estimer de manière non-paramétrique le modèle dynamique de l'état à partir d'observations bruitées. Enfin nous montrerons que cette approche permet en particulier de calculer la vraisemblance d'une séquence d'observations (model evidence) et ainsi trouver le "meilleur" catalogue pour représenter ces observations parmi une collection de catalogues (ces catalogues pouvant par exemple correspondre à des runs de plusieurs modèles).

### Références :

Andrieu, C., Doucet, A., & Holenstein, R. (2010). Particle markov chain monte carlo methods. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 72(3), 269-342.

Chau, T. T. T., Ailliot, P., Monbet, V., & Tandeo, P. (2018). An efficient particle-based method for maximum likelihood estimation in nonlinear state-space models. *arXiv preprint arXiv:1804.07483*.

Lguensat, R., Tandeo, P., Ailliot, P., Pulido, M., & Fablet, R. (2017). The analog data assimilation. *Monthly Weather Review*, 145(10), 4093-4107.

Lindsten, F., & Schön, T. B. (2013). Backward simulation methods for Monte Carlo statistical inference. *Foundations and Trends® in Machine Learning*, 6(1), 1-143.