

# Régression bayésienne pour la prédiction de quantiles spatio-temporels extrêmes

**Thomas OPITZ**, BioSP, INRA, France

**Raphaël Huser**, CEMSE Division, KAUST, Saudi Arabia

**Haakon Bakka**, CEMSE Division, KAUST, Saudi Arabia

**Håvard Rue**, CEMSE Division, KAUST, Saudi Arabia

**Mots-clés :** Integrated Nested Laplace Approximation, loi de Pareto généralisée, modélisation hiérarchique bayésienne, quantile extrême, théorie des valeurs extrêmes

La modélisation statistique de la queue de distribution permet de prédire la probabilité d'occurrence d'événements extrêmes à partir des données observées. La prise en compte de covariables explicatives peut considérablement améliorer ces prédictions. Nous proposons ici une approche de régression bayésienne pour modéliser des tendances complexes (ici, temporelles et spatiales) dans la queue de distribution, formulées à l'aide d'un système de modèles additifs généralisés pour la partie centrale de la distribution, pour la probabilité d'excès d'un seuil nonstationnaire élevé, et pour le dépassement au-dessus de ce seuil. Le modèle pour la partie centrale de distribution, estimé dans une première étape, nous permet de fixer un seuil nonstationnaire approprié.

Pour les dépassements de seuil, nous utilisons la loi de Pareto généralisée (GP), suggérée par les résultats asymptotiques en théorie des valeurs extrêmes. D'abord, nous dérivons la spécification de la "penalized complexity prior" pour l'indice de queue, un paramètre clé de la loi GP. Cette loi a priori concentre sa masse relativement proche d'un modèle de référence caractérisé par une rapide décroissance exponentielle dans la queue de distribution. Toutefois, les données peuvent contrebalancer une pénalité sur la distance de cette loi a priori par rapport à la loi exponentielle, permettant ainsi au modèle de capter des queues de distribution plus lourdes.

Dans les modèles de régression, nous modélisons des effets aléatoires latents, additifs et semi-paramétriques, à l'aide de processus gaussiens a priori, souples et faciles à interpréter. Nous nous appuyons sur INLA (Integrated Nested Laplace Approximation), une technique d'approximation analytique, pour estimer les lois a posteriori de façon rapide et précise.

Nous illustrons cette méthodologie en modélisant des tendances spatiales et saisonnières dans les quantiles extrêmes de données de cumuls de précipitations journalières aux Pays-Bas. En exploitant la rapidité de l'implémentation R-INLA, nous avons mené une étude extensive de validation croisée afin de sélectionner les valeurs de paramètres pilotant la régularité des courbes de tendance estimées. Nos résultats, présentés dans le cadre d'un challenge posé au sein de la conférence "Extreme Value Analysis 2017" à Delft (Pays-Bas), montrent une nette amélioration de prédiction en comparaison avec un simple benchmark, et sont comparables aux meilleurs scores d'évaluation des équipes participantes.

## Références

- [1] Opitz, T. and Huser, R. and Bakka, H. and Rue, H., (2018). INLA goes extreme: Bayesian tail regression for the estimation of high spatio-temporal quantiles. *Extremes*, To appear.