

# Projet de Thèse

## Sujet: Modélisation semi-paramétrique des extrêmes conditionnels

**Directeur de thèse:** ALIOU DIOP, Professeur Titulaire UFR SAT, UGB  
Saint-Louis

**Co-directeur de thèse:** STÉPHANE GIRARD, Directeur de Recherche INRIA,  
Grenoble, France

**Co-encadreur:** EL HADJI DEME, Enseignant-Chercheur UFR SAT, UGB  
Saint-Louis

### 1. DESCRIPTIF DÉTAILLÉ DU PROJET

**1.1. Problématique et objectifs.** Les événements extrêmes et catastrophiques (tremblements de terre, inondations, accidents nucléaires, crises monétaires ou financières, krachs boursiers, émergence d'un nouveau phénomène endémique tel que la maladie d'ébola) dominent l'actualité quotidienne par leur caractère imprévisible. Compte tenu de l'importance des enjeux sociaux et scientifiques, aucun débat sérieux sur le hasard ne saurait être mené sans une réflexion sur les événements rares et extrêmes.

La question que l'on pourrait se poser est de savoir ce que peuvent les statistiques face aux événements extrêmes ?

A cet effet, la théorie des valeurs extrêmes, apparue depuis quelques années, demeure un outil incontournable dans la compréhension et la gestion de ces phénomènes. Elle tente de les prévoir notamment par l'analyse des valeurs extrêmes d'une série de données complètes, censurées ou tronquées.

Par ce projet de recherche, nous tenterons d'adapter cette théorie afin de prendre en compte les composantes spatiales des phénomènes extrêmes (essentielles par exemple en pluviométrie, hydrologie ou épidémiologie) et d'apporter ainsi de nouveaux outils d'aide à la décision pour une meilleure gestion de ces phénomènes extrêmes.

Ce projet s'inscrit dans le cadre du projet SIMERGE qui est un projet d'équipe LIRIMA démarré en Janvier 2015. Voir <http://mistis.inrialpes.fr/simerge/>. Les partenaires du projet sont des chercheurs de l'équipe Mistis (Inria Grenoble - Rhône-Alpes, France), LERSTAD (Laboratoire d'Etudes et de Recherches en Statistiques

et Développement, Université Gaston Berger, Sénégal), IRD (Institut de Recherche pour le Développement, Unité de Recherche sur les Maladies Infectieuses et Tropicales Emergentes, Dakar, Sénégal) and LEM lab (Lille Economie et Management, Université Lille 1, 2, 3, Modal, Inria Lille Nord-Europe, France).

**1.2. Un nouveau modèle de queue de distribution.** De manière plus formelle, on s'intéresse à un couple  $(X, Y)$  de variables aléatoires où  $X$  est une covariable multivariée (spatiale) et  $Y$  une variable d'intérêt dépendant de la covariable  $X$  sous la forme  $Y = a(X) + b(X)Z$  où  $Z$  est une variable aléatoire indépendante de  $X$ . Les fonctions  $a(\cdot)$  et  $b(\cdot) > 0$  ne sont pas supposées connues et on ne fait pas d'hypothèse sur la loi de  $X$ . Pour modéliser  $Z$ , on s'intéressera à la classe des fonctions de répartition à variations régulières à l'infini. Cette classe est adaptée pour la modélisation de certains phénomènes catastrophiques ou extrêmes à queues lourdes. Ainsi, nous supposerons que la fonction de répartition de  $Z$  notée  $F$  appartient à cette classe et vérifie :

$$F(z) = 1 - z^{-1/\gamma} \ell(z), \quad \gamma > 0$$

où  $\ell$  est une fonction à variations lentes et  $\gamma$  est l'indice de queue. La fonction de répartition conditionnelle de  $Y$  est donnée par

$$\mathbb{P}(Y \leq y \mid X = x) = F\left(\frac{y - a(x)}{b(x)}\right) = 1 - \left(\frac{y - a(x)}{b(x)}\right)^{-1/\gamma} \ell\left(\frac{y - a(x)}{b(x)}\right).$$

On obtient ainsi un modèle semi-paramétrique de la loi de  $Y$  au point  $X = x$  de la covariable. Les fonctions  $a(\cdot)$  et  $b(\cdot)$  jouent le rôle de paramètres de position et d'échelle et dépendent de la covariable. L'indice de queue  $\gamma$  est lui supposé ne pas dépendre de la covariable, hypothèse courante dans les applications.

**1.3. Défis scientifiques.** Nous nous fixons comme objectifs de construire des estimateurs non-paramétriques pour les fonctions  $a(\cdot)$  et  $b(\cdot)$  de position et d'échelle ainsi que pour le paramètre de queue  $\gamma$ . Cette approche apporte l'avantage de plus de souplesse que les approches purement paramétriques, telle que celle introduite dans [9]. De plus, en supposant l'indice  $\gamma$  constant, nous espérons gagner en stabilité et robustesse vis à vis des approches purement non-paramétriques [4, 5, 6, 7, 8]. La modélisation semi-paramétrique que nous préconisons dans cette thèse nous paraît être ainsi un bon compromis entre généralité et robustesse. La construction d'estimateurs adaptés à cette nouvelle situation sera au centre de cette thèse. La difficulté sera d'obtenir les vitesses de convergence optimales dans ce contexte.

1.4. **Résultats attendus-Valorisation.** Nous espérons au terme de cette recherche:

- Etablir les propriétés asymptotiques des estimateurs proposés;
- Etendre l'étude au cas de données censurées ou tronquées en s'inspirant de [1, 2, 3];
- Appliquer les résultats obtenus sur des données réelles afin d'analyser un grand nombre de phénomènes extrêmes liés à la pluviométrie, l'hydrologie, la finance, l'épidémiologie etc;

Tous les résultats obtenus seront diffusés dans les revues scientifiques spécialisées, les universités et centres de recherches des pays concernés par ce projet de recherche et internet.

## 2. PROFILS DU DOCTORANTS

Le candidat doit être titulaire d'un master en Mathématiques Appliquées: option statistique. Des compétences en programmation sous R sont requises.

## 3. IDENTITÉ DU RESPONSABLE SCIENTIFIQUE ET LABORATOIRE DE RATTACHEMENT

Pr Aliou DIOP, Laboratoire d'Etude et de Recherche en Statistique et Développement,  
Université Gaston Berger.

## REFERENCES

- [1] Beirlant, J., Guillou, A., Toulemonde, G. Peaks-over-threshold modeling under random censoring. *Communication in Statistics-Theory and Methods*, 39:1158-1179, 2010.
- [2] Dabrowska, D. M. Nonparametric regression with censored survival time data. *Scandinavian Journal of Statistics*, 14:181-197, 1987.
- [3] Einmahl, J. H. J., Fils-Villetard, A., Guillou, A. Statistics of extremes under random censoring. *Bernoulli*, 14:207-227, 2008.
- [4] Gardes, L., Girard, S., 2008. A moving window approach for nonparametric estimation of the conditional tail index. *J. Multivariate Anal.* 99(10), 2368-2388.
- [5] Gardes, L., Girard, S., 2010. Conditional extremes from heavy-tailed distributions: an application to the estimation of extreme rainfall return levels. *Extremes* 13 (2), 177-204.
- [6] Gardes, L., Girard, S., Lekina, A., 2010. Functional nonparametric estimation of conditional extreme quantiles. *J. Multivariate Anal.* 101 (2), 419-433.
- [7] Goegebeur, Y., Guillou, A., Osmann, M. A local moment type estimator for the extreme value index in regression with random covariates. *The Canadian Journal of Statistics*, 42:487-507, 2014.

- [8] Goegebeur, Y., Guillou, A., Schorgen, A. Nonparametric regression estimation of conditional tails: the random covariate case. *Statistics*, 48:732-755, 2014.
- [9] Jalbert, J. Developing a non-stationary and regional frequency analysis method for intense precipitation events generated by the Canadian Regional Climate Model, Thèse de l'Université de Grenoble, 2014.