HAL open science

Modélisation des dépassements de seuils pour un processus observé à des temps irréguliers

Nicolas Raillard

To cite this version:

Nicolas Raillard. Modélisation des dépassements de seuils pour un processus observé à des temps irréguliers. 42èmes Journées de Statistique, 2010, Marseille, France, France. inria-00494835

HAL Id: inria-00494835

https://inria.hal.science/inria-00494835v1

Submitted on 24 Jun 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Modélisation des dépassements de seuils pour un processus observé à des temps irréguliers

Nicolas Raillard

IRMAR, UMR 6625 du CNRS, Université de Rennes 1, Rennes
Laboratoire d'Océanographie Spatiale, IFREMER, Brest
Laboratoire de Mathématiques de Brest, UMR 6205 du CNRS, UBO, Brest

Résumé

La modélisation des évènements extrêmes est de première importance pour garantir la pérennité des constructions humaines, et de manière plus générale pour assurer la continuité des activités humaines, tâche souvent dévolue à la théorie des valeurs extrêmes qui a pour but, entre autres, d'estimer la probabilité d'événements rarement observés, notamment celles de dépassement de seuils élevés. L'objectif de cette étude est d'estimer des telles quantités pour des phénomènes naturels échantillonnés de façon irrégulière.

Abstract

Modeling extremal events has become a main concern for practitioners to design structures, and to prevent interruptions of human activities : this task is often devoted to the extreme value theory, which is a convenient tool to model rare events, such as the probability that some process exceed a given threshold. The goal of the present study is to estimate such quantities for physical phenomenon, given an irregular time sampling.

De nombreuses activités et constructions humaines sont sensibles aux évènement extrêmes : c'est de ce constat qu'est née la théorie des valeurs extrêmes qui a pour objectif d'estimer la probabilité de certains événements rarement observés, qui peuvent être autant naturels — comme une tempête ou une vague de chaleur — que liés à l'activité humaine — comme une chute de cours de bourse.

L'objectif de cette étude est d'estimer la probabilité de dépasser des niveaux extrêmes pour des processus observées à temps irréguliers comme c'est souvent le cas en météorologie à cause de la présence de données manquantes (défaillance des capteurs, problèmes de transmission,...) ou de la nature des mesures (par exemple dans le cas de satellites, l'effet combiné de la rotation de la terre, du satellite, et de l'augmentation de leur nombre induit un échantillonnage complexe). Ceci se justifie d'autant plus dans un contexte d'observations spatialement réparties : il est bien rare d'avoir un réseau de capteurs formant une grille régulière dans l'espace (concentrations à certains endroits, zones moins bien échantillonnées...)

Pour des processus observés à temps régulier, deux types d'approches sont généralement utilisées (cf par exemple Coles (2001)) :

- **Maxima par blocs :** on regroupe les données par blocs de taille fixée, et on modélise la loi des maxima sur chaque bloc pour une loi des **valeurs extrêmes généralisée** (GEV). Il faut alors trouver un compromis entre la taille des blocs, qui doivent être assez long pour que l'approximation par la loi GEV soit réaliste, et le nombre de blocs, qui doit être assez grand pour obtenir une estimation précise des trois paramètres de la loi GEV. En pratique, le choix retenu est souvent de considérer les maxima annuels pour s'affranchir des effets saisonniers, ce qui nécessite d'observer le phénomène sur de nombreuses années. Cette méthode présente l'inconvénient bien connu de ne pas prendre en compte toute les informations présentes dans les données sur les événements extrêmes puisqu'elle ne conserve qu'un valeur par bloc pour l'inférence statistique. Une autre problème dans le cas de l'échantillonnage irrégulier est que la taille des blocs n'est plus constante : les paramètres de la loi GEV sont alors différents d'un bloc à l'autre.
- **Dépassements de seuil (POT) :** cette fois-ci tous les dépassements d'un seuil élevé sont modélisés, ce qui permet de conserver plus d'information que dans la méthode des maxima par bloc. Dans ce cas là, un des problèmes est celui de la dépendance temporelle entre deux observations successives, ce qui implique la survenue de groupes, ou clusters d'extrêmes. Deux solutions sont couramment utilisés : conserver uniquement une observation par groupe et supposer que les valeurs dans les groupes successifs sont indépendantes, ou bien prendre en compte la dépendance temporelle à l'intérieur des groupes. C'est dans ce cadre que nous allons nous placer.

Afin de mettre en place la solution évoquée ci-dessus, nous proposons un modèle permettant de décrire de manière réaliste la dépendance temporelle au dessus d'un niveau élevé dans le cas d'un échantillonage irrégulier. Cette approche, initiée par Smith et al. (1997) puis reprise par Coles et al. (2006) et Ribatet et al. (2009), conduit à la construction d'une fonction de vraisemblance qui prend en compte la censure des observations au dessous du seuil et la structure de dépendance temporelle dans les clusters. Dans un premier temps, nous étudierons la performance de cette méthode à l'aide de simulations avant de l'illustrer sur des données réelles (satellite ou bouée).

Références

[1] Stuart Coles. *An introduction to statistical modeling of extreme values.* Springer Series in Statistics. Springer-Verlag London Ltd., London, 2001.

[2] Stuart G. Coles, Jonathan A. Tawn, and Richard L. Smith. A seasonal markov model for extremely low temperatures. *Environmetrics*, Volume 5 Issue 3 :221 – 239, 2006.

[3] M. Ribatet, T.B. Ouarda, E. Sauquet, and J.M. Gresillon. Modeling All Exceedances Above a Threshold Using an Extremal Dependence Structure : Inferences on Several Flood Characteristics. *Water Resources Research*, 45 :W03407 p., 2009.

[4] Richard L. Smith, Jonathan A. Tawn, and Stuart G. Coles. Markov chain models for threshold exceedances. *Biometrika*, 84(2) :249–268, 1997.