Modélisation spatio-temporelle et précipitations extrêmes

J.N. Bacro[1] & C. Gaetan[2] & T. Opitz[3] & G. Toulemonde [1,4]

[1] *IMAG, Université de Montpellier, CNRS, Montpellier, France, jean-noel.bacro@umontpellier.fr*
[2] *DAIS, Università Ca' Foscari di Venezia, Venice, Italy, gaetan@unive.it*
[3] *BioSP, INRA, Avignon, France, thomas.opitz@inra.fr*
[4] *INRIA, Project-team LEMON, gwladys.toulemonde@umontpellier.fr*

Résumé. La modélisation statistique de données spatio-temporelles s'appuie aujourd'hui fortement sur des approches dites hiérarchiques. Ces dernières offrent en effet un cadre théorique relativement simple permettant une décomposition naturelle et exploitable des dépendances complexes inhérentes à ce type de données. Après avoir donné quelques éléments fondamentaux sur la modélisation spatio-temporelle, la présentation s'orientera sur des problématiques spécifiques aux données extrêmes. Un modèle spatio-temporel pour les dépassements de seuils élevés sera proposé et une application sur des précipitations dans le sud de la France sera présentée. L'intérêt et les limites de ce type de modélisation pour la simulation de scénarii de précipitations seront discutés.

Mots-clés. Extrêmes spatio-temporels, indépendance asymptotique, modèles hiérarchiques, champs aléatoires Gamma, convolution spatio-temporelle, vraisemblance composite.

Abstract. Statistical modeling of spatio-temporal data often relies heavily on so-called hierarchical approaches. The latter offer a relatively simple theoretical framework for exploitable decomposition of complex dependencies which are inherent to this type of data. After giving some fundamental elements on spatio-temporal modeling, the presentation will focus on issues specific to extreme data. A spatio-temporal model for high threshold exceedances will be proposed and an application on rainfall in the south of France will be presented. The interest and the limits of this type of modeling for simulation of rainfall scenarios will be discussed.

Keywords. Space-time extremes, asymptotic independence, hierarchical models, gamma random fields, space-time convolution, composite likelihood.

1 La modélisation des données spatio-temporelles

L'expression *données spatio-temporelles* fait allusion à des mesures auxquelles sont attachées la localisation et l'instant d'observation. De par leur nature, de telles données peuvent présenter des dépendances dans l'espace et/ou dans le temps, et ces dernières doivent être prises en compte dans les modélisations associées.

1.1 L'approche hiérarchique

La modélisation hiérarchique offre la possibilité de représenter des incertitudes provenant de différentes sources au travers de lois conditionnelles définies selon différents niveaux de hiérarchie. Cela permet de décomposer de façon assez naturelle des dépendances complexes. Une représentation classique distingue 2 niveaux de hiérarchie (Berliner, 1996): (i) *data model*, modélisation des données conditionnellement à un processus latent particulier et ses paramètres, (ii) *process model*, modélisation de l'incertitude attachée au processus caché conditionnellement aux valeurs de ses paramètres. L'approche bayesienne ajoute un niveau supplémentaire (iii), souvent appelé *parameter model*, où une loi jointe du vecteur des paramètres considérés est spécifiée. L'intérêt fondamental de l'approche hiérarchique est de faire apparaitre explicitement le processus latent que l'on suppose être sous-jacent aux observations obtenues.

1.2 Extrêmes de processus spatio-temporels

La modélisation des extrêmes de processus spatiaux a connu récemment des développements importants et les approches, initialement concentrées sur les maxima de réalisations, se sont peu à peu orientées sur les dépassements de seuils élevés. Cela permet, non seulement d'exploiter un plus grand nombre de données mais surtout de travailler à l'échelle des évènements extrêmes, ce qui facilite par exemple les interprétations physiques des phénomènes. S'appuyant sur la propriété de max-stabilité caractéristique des distributions extrêmes, différents modèles spatiaux ont été proposés, fondés sur des réalisations maximales ou sur des dépassements de seuils élevés, mais la propriété de stabilité de la dépendance des fortes valeurs qu'ils imposent peut s'avérer peu adaptée dans les applications. Une particularité des dépendances extrêmes réside en effet dans le fait que cette dépendance peut être soit constante à partir d'un certain seuil soit se réduire au fur et à mesure que l'on s'intéresse à des valeurs de plus en plus extrêmes. Cette dernière notion, appelée indépendance asymptotique des extrêmes, se vérifie par exemple pour les processus gaussiens et semble correspondre en pratique aux comportements des extrêmes de certains processus environnementaux. Cela impose donc de proposer des modèles alternatifs aux approches max-stables. Dans un cadre hiérarchique, différents modèles spatiaux ou spatio-temporels pour l'indépendance asymptotique ont été récemment proposés dans la littérature, essentiellement fondés sur des processus latents gaussiens et imposant de fait des dépendances extrêmes d'un type particulier (Casson and Coles, 1999; Gaetan and Grigoletto, 2007; Cooley et al., 2007; Sang and Gelfand, 2009). Dans la suite, nous proposons une modélisation hiérarchique pour les extrêmes de processus spatio-temporels, dans un cadre d'indépendance asymptotique, fondée sur un processus latent de type Gamma. L'approche, généralisant celle proposée dans un cadre temporel par Bortot et Gaetan (Bortot and Gaetan, 2014), assure une totale cohérence avec les résultats théoriques de la théorie des valeurs extrêmes pour le comportement des dépassements de seuils et offre la possibilité d'interprétation physique des évènements extrêmes réalisés dans le temps et l'espace.

2 Modélisation spatio-temporelle de dépassements de seuils élevés

2.1 Le modèle proposé

Soit X une variable aléatoire continue dont la distribution appartient à un domaine d'attraction de maximum. Asymptotiquement, la distribution des dépassements de seuils élevés de X peut être modélisée par une distribution de Pareto généralisée (PG), donnée par $PG(y;\sigma,\xi) = 1 - \left(1+\xi\frac{y}{\sigma}\right)_+^{-(1/\xi)}$ où $\xi \in R$ est un paramètre de forme caractéristique du domaine d'attraction du maximum et σ un paramètre d'échelle positif. Lorsque $\xi > 0$, la distribution PG peut être obtenue comme un mélange Gamma - Exponentiel (Reiss and Thomas, 2007, p.157) : ,

$$V1|V2 \sim \mathrm{Exp}(V2), \quad V2 \sim \mathrm{Gamma}(1/\xi, \sigma/\xi) \quad \Rightarrow \quad V1 \sim PG(\,\cdot\,;\sigma,\xi),$$

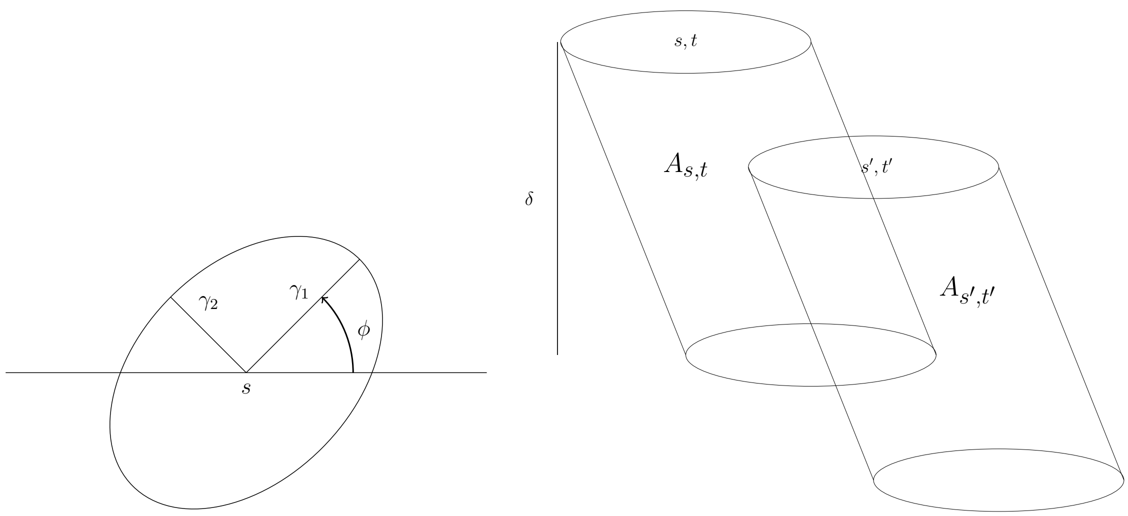
où $\mathrm{Exp}(b)$ représente la loi Exponentielle de paramètre $b > 0$ et $\mathrm{Gamma}(a,b)$ la loi Gamma de paramètres (a,b). L'exploitation de cette structure hiérarchique nous permet de proposer, dans l'esprit de Bortot and Gaetan (2014), un modèle spatio-temporel stationnaire pour les dépassements d'un seuil élevé (Bacro et al., 2019). La construction est la suivante : soit $Z = \{Z(x), x \in \mathcal{X}\}$ un processus spatio-temporel, avec $x = (s,t)$ et $\mathcal{X} = \mathbb{R}^2 \times \mathbb{R}^+$, où s représente la localisation et t le temps. Sans perte de généralité, $Z(x)$ est supposé appartenir au domaine d'attraction de Fréchet ($\xi > 0$). Soit $Y(\cdot) = (Z(\cdot) - u)\mathbf{1}_{(u,\infty)}(Z(\cdot))$ le processus des dépassements d'un seuil u associé à Z et $\{\Lambda(x), x \in \mathcal{X}\}$ un processus spatio-temporel latent de distributions marginales $\Lambda(x) \sim \mathrm{Gamma}(\alpha,\beta)$. On considère alors le modèle hiérarchique suivant :

$$Y(x) \mid [\Lambda(x), Y(x) > 0] \sim \mathrm{Exp}\left(\Lambda(x)\right), \tag{1a}$$

$$\Pr(Y(x) > 0 \mid \Lambda(x)) = e^{-\kappa\Lambda(x)}, \tag{1b}$$

où $\kappa > 0$ est un paramètre contrôlant le taux de dépassements de u. On a alors $\Pr(Z(x) > u) = \mathrm{E}\left(\Pr(Y(x) > 0|\Lambda(x))\right) = \mathrm{E}\left(e^{-\kappa\Lambda(x)}\right) = \left(\frac{\beta}{\kappa+\beta}\right)^{\alpha}$.

Le second niveau de la modélisation consiste à spécifier la dépendance spatio-temporelle portée par le processus $\{\Lambda(x), x \in \mathcal{X}\}$. Soit $\Gamma(\cdot)$ un champ Gamma défini comme une mesure non négative sur $\mathcal{X}$ caractérisée par une mesure $\alpha(dx)$ et un paramètre β, tels que $\forall B \in \mathcal{B}_b(\mathcal{X})$, $\Gamma(B) := \int_B \Gamma(dx) \sim \mathrm{Gamma}(\alpha(B), \beta)$, avec $\alpha(B) = \int_B \alpha(dx)$ (Ferguson, 1973). L'idée est alors d'exploiter une approche par convolution comme proposée dans Wolpert and Ickstadt (1998): $\Lambda(x) = \int K(x,x')\Gamma(dx')$, où $K(x,x') = \mathbf{1}_A(x-x')$ est la



(a) (b)

Figure 1: Noyaux spatio-temporels. A gauche : ellipse spatiale $E(s, \gamma_1, \gamma_2, \phi)$ centrée en s. A droite : intersection entre deux cylindres $A_{s,t}$ et $A_{s',t'}$, centrés en (s, t) et (s', t'), évènement de durée δ.

fonction indicatrice d'un cylindre A, éventuellement incliné, dont la base est une ellipse. Une interprétation physique possible est alors la suivante : l'ellipse décrit la zone spatiale d'influence d'un évènement extrême (*tempête*) centré en s ; la hauteur du cylindre est associée à la durée de la tempête et le caractère plus ou moins incliné du cylindre reflète la dynamique de déplacement. Une représentation simple de ces notions et des paramètres qui y sont associés est donnée Figure 1. On peut montrer que le modèle proposé est un modèle présentant de l'indépendance asymptotique. L'inférence statistique des paramètres du modèle est réalisée selon une approche vraisemblance composite par paires.

2.2 Application : précipitations horaires dans le sud de la France

Le modèle proposé est appliqué sur un jeu de données de précipitations horaires mesurées en 50 stations du sud de la France, sur la période septembre-octobre des années 1993 à 2014 (54542 heures). La configuration spatiale des données est illustrée Figure 2. L'indépendance asymptotique des données est validée au travers d'une étude exploratoire. Dans un but de comparaison, cinq modèles correspondant au cadre d'indépendance asymptotique sont ajustés sur les données (Bortot et al., 2000; Renard and Lang, 2007; Davison et al., 2013) : le modèle proposé ci-dessus, avec et sans vitesse ; un modèle séparable fondé sur une copule bivariée gaussienne ; deux modèles de type frozen, avec et sans support compact (voir Christakos, 2017, pour une présentation détaillée).

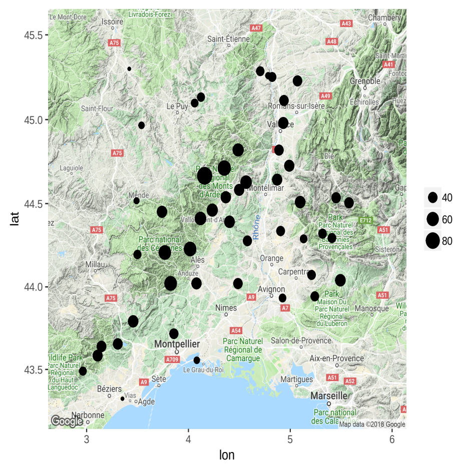


Figure 2: Données de précipitations sur le sud de la France. Le diamètre des points est proportionnel au quantile empirique à 99%.

Une comparaison de ces différentes modélisations, s'appuyant sur les valeurs des CLIC associés à chaque modèle et sur des représentations graphiques de probabilités conditionnelles extrêmes estimées et empiriques selon différentes directions spatiales et temporelles, permet de montrer que le modèle spatio-temporel de dépassements proposé avec vitesse non nulle offre le meilleur ajustement des données considérées.

References

Bacro, J.-N., Gaetan, C., Opitz, T., and Toulemonde, G. (2019), "Hierarchical space-time modeling of asymptotically independent exceedances with an application to precipitation data," *Journal of the American Statistical Association*, to appear.

Berliner, L. M. (1996), "Hierarchical Bayesian time-series models," in *Maximum Entropy and Bayesian Methods*, Dordrecht: Kluwer Academic, pp. 15–22.

Bortot, P., Coles, S., and Tawn, J. (2000), "The multivariate Gaussian tail model: an application to oceanographic data," *Journal of the Royal Statistical Society: Series C*, 49, 31–49.

Bortot, P., and Gaetan, C. (2014), "A latent process model for temporal extremes," *Scandinavian Journal of Statistics*, 41, 606–621.

Casson, E., and Coles, S. G. (1999), "Spatial regression models for extremes," *Extremes*, 1, 449–468.

Christakos, G. (2017), *Spatiotemporal Random Fields Theory and Applications*, Amsterdam: Elsevier.

Cooley, D., Nychka, D., and Naveau, P. (2007), "Bayesian Spatial Modeling of Extreme Precipitation Return Levels," *Journal of the American Statistical Association*, 102, 824–840.

Davison, A. C., Huser, R., and Thibaud, E. (2013), "Geostatistics of dependent and asymptotically independent extremes," *Journal of Mathematical Geosciences*, 45, 511–529.

Ferguson, T. S. (1973), "A Bayesian Analysis of Some Nonparametric Problems," *The Annals of Statistics*, 1, 209–230.

Gaetan, C., and Grigoletto, M. (2007), "A hierarchical model for the analysis of spatial rainfall extremes," *Journal of Agricultural Biological and Environmental Statistics*, 12, 434–449.

Reiss, R., and Thomas, M. (2007), *Statistical Analysis of Extreme Values*, third edn, Basel: Birkhäuser.

Renard, B., and Lang, M. (2007), "Use of a Gaussian copula for multivariate extreme value analysis: Some case studies in hydrology," *Advances in Water Resources*, 30, 897 – 912.

Sang, H., and Gelfand, A. (2009), "Hierarchical modeling for extreme values observed over space and time," *Environmental and Ecological Statistics*, 16, 407–426.

Wolpert, R. L., and Ickstadt, K. (1998), "Poisson/Gamma random fields for spatial statistics," *Biometrika*, 85, 251–267.