

SIMULATION CONDITIONNELLE DE PROCESSUS MAX-STABLES POUR LE DOWNSCALING DES EXTRÊMES

Aurélien Bechler^{1,2,3} & Liliane Bel^{1,2} & Mathieu Vrac³

¹ INRA, UMR UMR 518 Math. Info. Appli., F-75005 Paris, France et
aurelien.bechler@agroparistech.fr

² AgroParisTech, UMR 518 Math. Info. Appli., F-75005 Paris, France et
liliane.bel@agroparistech.fr

³ LSCE-IPSL, Centre d'Etudes de Saclay, Orme des Merisiers, Bat. 701 91191 Gif-sur-Yvette,
France et *mathieu.vrac@lsce.ipsl.fr*

Résumé. Depuis peu, les modèles régionaux de climat permettent de fournir des scénarios futurs de précipitation à une échelle de plus en plus précise. Néanmoins, cette résolution n'est pas encore suffisante pour décrire précisément ce qu'il se passe à une échelle ponctuelle. Des méthodes de descente d'échelle, ou downscaling (Maraun et al., 2010), ont été développées et permettent de faire le lien entre les deux niveaux de résolution. Un problème apparaît lorsque l'on s'intéresse au comportement extrême de ces phénomènes qui est mal décrit par ces modèles. Pour le résoudre, nous proposons une méthodologie en deux étapes. La première consiste à construire une fonction de transfert entre la variable grande échelle et la variable petite échelle en certains points de la carte. Puis, une fois cette correction de biais effectuée, l'algorithme de simulation conditionnelle de processus max-stables proposée par Dombry et al. (2013) et adapté pour le processus t extremal (Bechler et al., 2014) permet d'obtenir les distributions des précipitations extrêmes en tout point de la carte. Cette méthode a été appliquée pour simuler des données de précipitations extrêmes à partir de scénarios grande échelle dans le sud de la France, région dans laquelle leurs conséquences (entre autres les inondations) peuvent être désastreuses.

Mots-clés. Simulation Conditionnelle, Extrêmes Spatiaux, Processus t Extremal, Downscaling.

Abstract. For a few years, regional climate models are used to provide future scenarios of precipitation with increasingly higher resolution. However, this resolution is not yet sufficient to describe efficiently what happens at local scale. Methods of downscaling (Maraun et al., 2010) have been developed and allow us to make the link between these two levels of resolution. Nevertheless, the extreme behavior is not well described by these downscaling methods. We propose a two step methodology to solve this problem. The first one consists in building a statistical link-function between the high scale and low-scale variables at some given locations. Once this bias correction performed, the conditional simulation algorithm of max-stable processes proposed by Dombry et al. (2013) and adapted to the extremal t process (Bechler et al., 2014) enables us to get extreme precipitation distributions at any point of the map. Application is performed on precipitation data in the south of France where extreme (evenol) events have major impact (e.g. floods).

Keywords. Conditional Simulation, Spatial Extremes, Extremal t Process, Downscaling.

1 Introduction

Les précipitations dans le sud de la France peuvent atteindre des niveaux très importants lors d'évènement appelés cévenols. Les modèles de la famille max-stable, qui permettent de décrire des processus de maxima avec des structures de dépendance dans les extrêmes, constituent un bon outil pour modéliser ce genre d'évènements. Le modèle t-extremal choisi dans ce travail est assez souple pour permettre de présenter une grande variété de situations de dépendance asymptotique. Pour appréhender les comportements des précipitations extrêmes en des lieux vides d'observations, une adaptation de l'algorithme proposé par Dombry et al. (2013) permet de réaliser des simulations conditionnelles de processus t-extremal (Bechler et al., 2014).

Dans un cadre de prévisions futures, on cherche à pouvoir utiliser cet algorithme lorsqu'il n'y a plus d'observations ponctuelles disponibles. Pour cela, on va utiliser de l'information grande échelle fournie par des modèle de climat régionaux (RCM). Ces sorties de modèle ont une résolution spatiale moyenne et des difficultés à modéliser les comportements extrêmes des phénomènes. Une première étape consistera donc à effectuer une correction de biais de ces variables pour obtenir des pseudo-observations le proche possible de la réalité. Une discussion sur la méthode de correction de biais sera détaillée. Par la suite, la procédure de simulation conditionnelle pourra être utilisée directement en conditionnant par rapport à ces pseudo-observations pour obtenir des distributions de précipitations extrêmes en tout point de la région.

2 Processus t extremal

Dans ce travail, le modèle qui a été choisi pour décrire les évènements extrêmes est celui du processus t extremal (Padoan, 2011) défini par l'équation

$$Z(x) = m_\alpha^{-1/\alpha} \max_{i \geq 1} \xi_i W_i(x), \text{ avec } x \in \mathbb{R}^2, \alpha > 0, \quad (1)$$

avec $m_\alpha = \sqrt{\pi}^{-1} 2^{(\alpha-2)/2} \Gamma((\alpha+1)/2)$, $\{\xi_i\}_{i \geq 1}$ un processus de Poisson sur $(0, +\infty)$ d'intensité $\alpha t^{-(\alpha+1)} dt$ et W_i des copies indépendantes d'un champ gaussien stationnaire sur \mathbb{R}^2 , centré et de fonction de corrélation ρ . Sous ces conditions, $Z(x)$ est un processus max-stable stationnaire de marginales α -Fréchet. Il apparaît que pour $\alpha = 1$, l'équation (1) décrit un modèle de Schlather (Schlather, 2002).

L'algorithme de simulation conditionnelle de Dombry et al. (2013), initialement utilisé pour les processus de Schlather et Brown-Resnick a été adapté pour le processus t extremal (Bechler et al., 2014).

3 Présentation des données

Données SAFRAN (haute résolution)

Les données utilisées dans ce travail sont issues des réanalyses SAFRAN (Quintana-Segui et al., 2008) qui couvrent la région sud-est de la France. On en a extrait les maxima automnaux des précipitations journalières de 1960 à 2007 sur une grille de 457 points uniformément répartis. La figure 1 montre l'emplacement géographique de ces données.

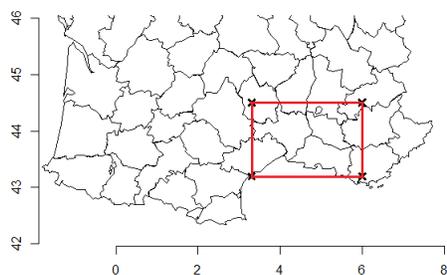


FIGURE 1 – Zone d'étude des données SAFRAN.

L'apprentissage sera effectué sur les années 1960-1999 et les tests sur la période 2000-2007. Les paramètres du modèle t extrémal (α , ρ ainsi que les paramètres GEV des marginales) sont inférés par maximisation de la vraisemblance complète (Thibaud et Opitz, 2013).

Données EURO-CORDEX (basse résolution)

Les données EURO-CORDEX sont issues des sorties de modèle RCM sur la période 1989-2007. On utilise les moyennes journalières sur une grille de résolution 50 km. Ainsi, les données couvrent notre région d'intérêt avec 15 mailles. Sur chaque maille, on construit le maximum automnal des précipitations moyennes journalières. La figure 2 montre ces maxima construits à partir des données EURO-CORDEX.

On considère les années 1989-1999 comme période d'apprentissage et les années 2000-2007 sont une nouvelle fois utilisées pour la vérification.

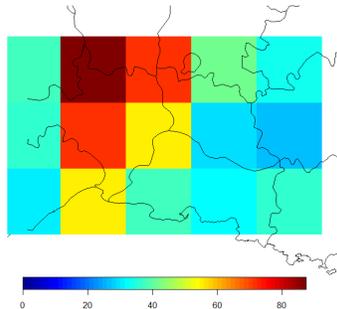


FIGURE 2 – Données EURO-CORDEX des maxima automnaux des précipitations pour l’année 1989 (en mm).

4 Calibration

Pour chaque maille de 50km des données EURO-CORDEX, on cherche parmi les points de la grille SAFRAN le point situé géographiquement le plus au centre. Ces points vont servir à établir un lien statistique entre les valeurs des RCM et les données SAFRAN. Ce lien, établi sur la période d’apprentissage, va permettre de corriger les sorties des RCM sur la période test, pour se rapprocher le plus possible du maximum observé. Les valeurs obtenues, appelées pseudo-observations, serviront de valeurs conditionnantes pour l’algorithme de simulation conditionnelle. Différentes techniques de calibration sont proposées par la suite.

Méthode CDF-t

La méthode CDF-transform consiste à construire une méthode de correction de biais en comparant les fonctions de répartition d’une variable grande échelle et celle d’une variable faible échelle (Michelangeli et al., 2009).

Régression linéaire

Une régression linéaire est utilisée pour modéliser le lien entre la variable grande échelle issue des RCM et la variable ponctuelle issue des données SAFRAN au point choisi.

5 Application

Critères de comparaison

L’efficacité des simulations conditionnelles est évaluée en utilisant le continuous ranking probability score (Gneiting et Raftery, 2007) crps. En notant F la fonction de répartition empi-

rique des simulations conditionnelles et y l'observation test, le crps peut s'écrire

$$\text{crps}(F; y) = \int_{-\infty}^{+\infty} [F(t) - H(t - y)]^2 dt,$$

avec $H(u)$ la fonction de Heavyside qui est égale à 0 pour $u < 0$ et 1 sinon.

De plus, on définit le quantile score (qs) comme

$$\text{qs}_p(F; y) = \nu_p(y - F^{-1}(p))$$

avec p la probabilité d'intérêt et $\nu_p(u) = pu$ si $u \geq 0$ et $\nu_p(u) = (p - 1)u$ sinon. En choisissant des probabilité p suffisamment élevées, le quantile score s'intéresse aux performances prédictives des méthodes dans la queue de distribution.

Des scores globaux $\overline{\text{CRPS}}$ et $\overline{\text{QS}}_p$ sont construits en moyennant sur tous les points test. Dans les deux cas, un bon score est un score proche de 0.

Résultats

Pour chaque méthode de calibration, on effectue des simulations conditionnelles aux 442 points de la grille à partir des 15 pseudo-observations. Différents modèles sont comparés. Le modèle RCM-Brut est obtenu en utilisant pour chaque maille directement les valeurs des RCM. Ainsi tous les points inclus dans la même maille auront la même valeur. Les modèles RCM-CDFt et RCM-RL utilisent respectivement la méthode CDF-t et la régression linéaire comme fonctions de transfert pour obtenir les valeurs conditionnantes. Puis l'algorithme de simulation conditionnelle est appliqué en utilisant ces valeurs conditionnantes. Pour finir, nous comparons ces résultats au modèle optimal qui prend directement les valeurs SAFRAN comme valeurs conditionnantes. Cela correspond au cas de figure où la fonction de transfert corrigerait de manière parfaite le biais des RCM. Le tableau 1 détaille les scores obtenus aux 442 points test pour les différentes modèles.

Modèles	$\overline{\text{CRPS}}$	$\overline{\text{QS}}_{95}$
RCM-Brut	32.12	27.05
RCM-CDFt	24.68	6.43
RCM-RL	19.93	7.17
Optimal	16.98	4.59

TABLE 1 – $\overline{\text{CRPS}}$ et $\overline{\text{QS}}_{95}$ de l'algorithme avec différentes méthodes pour construire les pseudo-observations.

Dans un premier temps, ces résultats montrent un vrai avantage à utiliser la simulation conditionnelle en terme de $\overline{\text{CRPS}}$ et $\overline{\text{QS}}_{95}$. De plus, l'utilisation de différentes fonctions de transfert influe directement sur la qualité des simulations obtenues. Ainsi, le modèle RCM-RL

est meilleur que RCM-CDFt en terme de $\overline{\text{CRPS}}$ mais moins bon en $\overline{\text{QS}}_{95}$. Dans tous les cas, il apparaît néanmoins que les scores obtenus sont encore relativement éloigné des scores du modèle optimal. Même si ce dernier restera inatteignable, d'autres fonctions de transfert aideraient peut-être à s'en rapprocher.

Bibliographie

- [1] Bechler A., Bel L., Vrac M., (2014), *Conditional simulations of the extremal t process : application to fields of extreme precipitation*, Soumis.
- [2] Dombry C., Eyi-Minko F, Ribatet M., (2013), *Conditional simulation of max-stable processes*, Biometrika, Vol. 100 (1), 111-124.
- [3] Gneiting T., Raftery A. E., 2007, *Strictly Proper Scoring Rules, Prediction, and Estimation*, Journal of the American Statistical Association, Vol. 102 (477), 359-378.
- [4] Maraun D., Wetterhall F., Ireson A. M., Chandler R. E., Kendon E. J., Widmann M., Brienen S., Rust H. W., Sauter T., Themeßl M., Venema V. K. C. Chun, K. P., Goodess C. M., Jones R. G., Onof C., Vrac M., Thiele-Eich I., 2010, *Precipitation downscaling under climate change : Recent developments to bridge the gap between dynamical models and the end user*, Reviews of Geophysics, Vol. 48 (3).
- [5] Michelangeli P.-A., Vrac M., Loukos H., 2009, *Probabilistic downscaling approaches : Application to wind cumulative distribution functions*, Geophysical Research Letters, Vol. 36 (11).
- [6] Padoan, S. A., 2011, *Multivariate Extreme Models Based on Underlying Skew- t and Skew-Normal Distributions*, Journal of Multivariate Analysis, Vol. 102 (5), 977-991.
- [7] Quintana-Segui P., Le Moigne P., Durand Y., Martin E., Habets F., Baillon M., Canellas C., Franchistéguy L., Morel, S., 2008, *Analysis of near surface atmospheric variables : validation of the SAFRAN analysis over France*, Journal of Applied Meteorology and Climatology, Vol. 47 (1), 92-107.
- [8] Thibaud E., Opitz T., (2013), *Efficient inference and simulation for elliptical Pareto processes*, arXiv :1401.0168v1.
- [9] Schlather, M. (2002), *Model for Stationary Max-Stable Random Fields*, Extremes 5 :1, 33-44.