

# Applications de l'approche bayesienne à la représentation des pluies mensuelles africaines par la loi des fuites

Jacques BERNIER, Eric PARENT et Jean-Jacques BOREUX  
(Laboratoire de Gestion des Risques En Sciences de l'Environnement),  
ENGREF, 19, Avenue du Maine, F-75732 Paris Cedex 15,  
Parent@engref.fr, Jacques.Bernier2@wanadoo.fr, Boreux@fuc.ac.be

*GESTION DU RISQUE EAU EN PAYS SEMI-ARIDE*  
Conférence de la Société Hydrotechnique de France  
21 et 22 mai 2003, TUNIS

**Résumé :** Ce texte présente une analyse bayesienne d'un processus ponctuels marqué particulier (la loi des fuites) avec application à des problèmes de modélisation environnementale hydrométéorologique sur les données Africaines de Ghezala (Tunisie) et Ataktamé (Togo).

**Mots-clès :** modèle bayesien, méthodes Monté Carlo par Chaines de Markov, processus ponctuel marqué

**Abstract :** This text presents a Bayesian analysis of a specific marked point process (the law of the leaks) with application for hydrometeorological environmental modelling on the African data of Ghezala (Tunisia) and Ataktamé (Togo).

**key words :** Bayesian models ; Markov Chain Monte Carlo methods ; Marked Point Process

# 1 Introduction

La modélisation des " distributions de fréquences des précipitations " est un thème fort ancien de l'hydrométéorologie. Il est encore d'actualité et surtout d'intérêt vital sur le continent africain notamment lorsque ses résultats peuvent être appliqués à la prédition des déficits et sécheresses (Tapsoba, 1997). Il n'est pas dans notre propos de faire un inventaire exhaustif des modèles disponibles pour la représentations des pluies mensuelles (cf Handbook of Hydrology de Maidment, 1993, par exemple). La plupart des modèles proposés ont des difficultés à bien représenter les fréquences de pluies nulles ou faibles significativement présentes dans les contrées arides et semi arides. En son temps nous avions proposé la distribution dite " loi des fuites " (Bernier, 1970 ; Bernier, Fandeux - 1970). Ce modèle de distribution mixte a été inclus dans le logiciel Hyfran (2000) de la Chaire en Hydrologie de l'université du Québec. Malgré cela son usage semble avoir été limité par suite des difficultés de calcul inhérentes à l'approche statistique fréquentiste utilisée pour apprécier l'incertitude des estimations notamment.

Nous reprenons ici ce problème dans le cadre de l'approche bayesienne combinée avec les méthodes de calcul modernes dites MCMC (méthodes de Monte Carlo par chaînes de Markoff) décrites par exemple dans Robert et Casella (1998). L'approche Bayésienne, basée sur l'emploi des raisonnements conditionnels probabilistes (aussi bien au niveau de la modélisation qu'au niveau des techniques d'estimation) est particulièrement adaptée aux problèmes d'environnement (Bernier et al, 2000). Elle est de plus complète car on peut l'appliquer en prédition notamment en vue de gestion de ressources futures en y incluant précisément l'ensemble des incertitudes analysées et quantifiées dans la première phase d'estimation du modèle.

## 1.1 Les modèles stochastiques ponctuels marqués

La figure 1 présente la trajectoire théorique d'un processus ponctuels marqué où des événements ponctuels (supposés sans durée) surviennent aléatoirement dans le temps.

A chaque occurrence d'un événement est associée une marque  $X_i$ , variable aléatoire que l'on considère au dessus d'un seuil  $u_0$ . Le modèle le plus simple de ce type suppose des événements indépendants survenant dans le temps selon un processus de Poisson tel que :

- si  $N_T$  est le nombre d'événements sur un intervalle fixé  $[0, T]$  , ce nombre

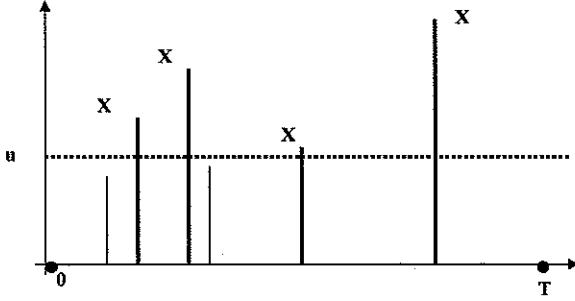


FIG. 1 – Un exemple de trajectoire de processus ponctuel marqué

est distribué selon la " loi de Poisson " :

$$[N_T = n_j | \mu, T] = e^{\mu T} \frac{(\mu T)^{n_j}}{n_j!}$$

Nous adoptons ici la notation commode  $[N_T = n_j | \mu, T]$  d'une densité de probabilité d'une variable, ici  $N_T$ , où, à droite de la barre  $|$ , sont inscrits les variables ou paramètres de conditionnement.

- les  $X_i$  successifs sont indépendants et leurs dépassemens d'un seuil  $u_0$  sont distribués selon la " loi exponentielle " :

$$[X_i \leq x | \rho, X_i \geq u_0] = 1 - e^{-\rho(x-u_0)}$$

- pour  $u_0$  fixé, les deux paramètres  $\mu$  et  $\rho$  caractérisent entièrement le modèle.

Ce modèle a de multiples applications dans le domaine de l'environnement concernant notamment les valeurs extrêmes (Coles, 2001). En hydrologie (Wang, 1991) on connaît le modèle POT où l'on s'intéresse aux crues ou aux pluies extrêmes dépassant un niveau  $u_0$ . Il s'agit alors de considérer des variables associées du type :

$$Y_T = \max_{[0, T]} (X_i)$$

Cependant on peut s'intéresser à d'autres variables associées à ce processus comme les sommes :

$$P_T = \text{Somme}_{[0,T]}(X_i - u_0) = \sum_{i=1}^{N_T} (X_i - u_0)$$

Dans le contexte des extrêmes, il peut s'agir, par exemple, de " dommages cumulés " générés par les crues supérieures au seuil. Dans cette communication nous nous intéresserons aux précipitations totales mensuelles ( $T=1$  mois, avec  $u_0 = 0$ ) considérées comme sommes de pluies ponctuelles instantanées  $X_i$  décrites par le processus ci dessus. Si aux très courtes échelles de temps, ce modèle n'est pas réaliste, c'est surtout l'hypothèse de pluie instantanée et non pas la distribution des intensités qui est en cause. Cependant à une échelle de temps comme le mois, la chronologie des pluies importe moins. Ceci suggère l'emploi de ce modèle avec  $u_0 = 0$  pour les pluies mensuelles.

## 1.2 La " loi des fuites "

Un calcul mathématique simple donne l'expression de la distribution de  $P_T$  sous la forme :

$$\begin{aligned} [P_T | \mu, \rho] &= \mu \rho T \cdot \exp(-\mu T - \rho P_T) \frac{I_1(2\sqrt{\mu \rho T P_T})}{\sqrt{\mu \rho T P_T}} \quad \text{si } P_T > 0 \quad (1) \\ [P_T = 0] &= \exp(-\mu T) \end{aligned}$$

On notera qu'il s'agit d'une distribution mixte avec probabilité finie non nulle à l'origine. Dans cette expression apparaît une fonction de Bessel modifiée de seconde espèce  $I_1$ . Nous l'appelons " loi des fuites " selon la terminologie de G. Morlat qui a suggéré le modèle pour la représentation du volume de fuites sur les canalisations de transport de gaz. A notre connaissance les premières applications à l'hydrologie sont dues à Fandeux et Bernier (1977). Le logiciel Hyfran (2000) en facilite l'estimation selon les méthodes classiques avec toutefois des imprécisions et approximations qui peuvent en limiter l'usage pour des séries trop courtes.

Selon les principes bayésiens, toute l'information nécessaire à l'estimation du modèle est contenue dans la distribution conditionnelle a posteriori :

$$[\mu, \rho | P^{(N)}] = \frac{[P^{(N)} | \mu, \rho] \cdot [\mu, \rho]}{\int \int [P^{(N)} | \mu, \rho] \cdot [\mu, \rho] d\mu d\rho}$$

où  $[P^{(N)}|\mu, \rho] = \prod_{i=1}^{i=N} [P_i|\mu, \rho]$  est la vraisemblance de  $N$  pluies mensuelle indépendantes observées et représentées par le modèle (1) et  $[\mu, \rho]$  est la distribution a priori des paramètres soit non informative soit construite sur des jugements d'expert.

Les techniques d'estimation ne sont pas décrites dans cet article. Elles sont fondées sur l'emploi d'algorithmes Monte carlo par Chaines de Markov (MCMC), qui sont aujourd'hui les techniques standard pour l'estimation des modèles Bayésien. On trouve dans Kuczera et Parent (1998) les éléments d'initiation à leur emploi en hydrologie. L'algorithme employé ici (Tanner 1996) est analogue à celui décrit dans Parent et Bernier (2002).

## 2 Applications à deux stations Africaines

Nos illustrations, non exhaustives des possibilités de ce modèle utilisant l'approche bayesienne, concernent deux stations météorologiques africaines : la station Atakpamé au Togo (période utilisée avec lacunes de 1899 à 1987, soit un échantillon de  $N = 64$  ans) et la station Ghézala-Barrage en Tunisie (période utilisée de 1968 à 2001,  $N = 34$  ans). Les mois de Décembre et Juillet pour Atakpamé, Février et Août pour Ghézala - Barrage ont été choisis. Les distributions observées avec les histogrammes sur 20 classes des précipitations mensuelles pour ces deux sites en deux périodes sont présentées en figure 2.

On notera les différences de régimes avec des distributions en J comportant une fréquence de zeros significative en Février (Togo) et Août (Tunisie) et l'apparence de deux modes pour les pluies de Juillet à Atakpamé. Pour notre application, on notera aussi les différences de tailles d'échantillon entre les deux lieux. A titre d'exemple, la figure 3 concerne les densités de probabilité a posteriori des deux paramètres  $\mu$  et  $\rho$  pour la station de Ghezala en Aout. Les histogrammes (à droite) ont été obtenus par simulation MCMC. Les figures de gauche montrent les lissages efficaces obtenus par la technique dite de " Rao-Blackwell ". Cette technique, décrite dans Rivot et al (2000) est applicable aisément ici.

Les tableaux 1 et 2 donnent les valeurs moyennes et limites des intervalles de crédibilité a posteriori à 80% (limites inférieure et supérieures respectivement à 10 et 90%) des paramètres et de 3 caractéristiques des distributions mensuelles. Il s'agit des quantiles des pluies mensuelles : médiane, valeurs décennale et cinquantenale.

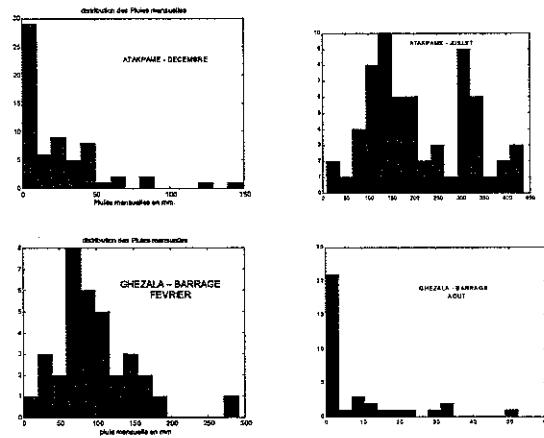


FIG. 2 – Histogrammes des données mensuelles de pluies à Atakpamé et à Ghezala

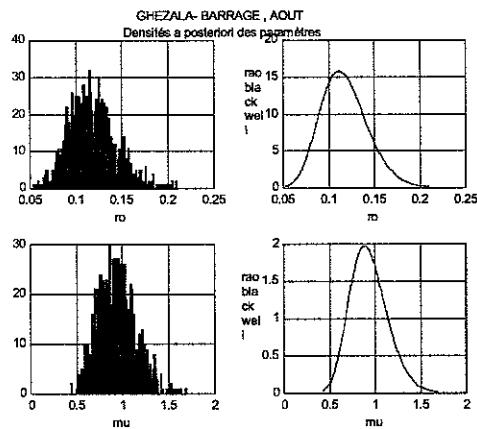


FIG. 3 – Distributions des paramètres obtenues par simulations MCMC

<i>ATAKPAME</i>	<i>Décembre</i>			<i>Juillet</i>		
$\mu$	1.14	0.87	1.47	7.37	5.53	9.69
$\rho$	0.0490	.037	0.062	0.035	0.026	0.046
Pluie médiane	9.3	4.0	15.9	194.5	172.4	216.9
Pluie décennale	63.9	49.3	81.6	359.8	322.4	403.7
Pluie cinquantenaire	113.9	89.3	143.9	483.3	426.8	549.4

**Tableau 1 : Intervalles de crédibilité des caractéristiques a posteriori de la station d'Atakpamé**

<i>GHEZALA</i>	<i>Février</i>			<i>Février</i>		
$\mu$	6.26	4.14	8.73	0.93	0.63	1.28
$\rho$	0.066	0.044	0.093	0.118	0.077	0.167
Pluie médiane	86.3	70.4	102.7	2.6	1.4	5.2
Pluie décennale	169.0	142.2	200.7	23.4	15.8	34.0
Pluie cinquantenaire	230.4	188.6	278.2	43.6	30.1	62.7

**Tableau 2 : Intervalles de crédibilité des caractéristiques a posteriori de la station de Gezhala**

On notera l'étendue des intervalles de crédibilité plus large pour Ghezala-Barrage est fonction de la quantité d'information plus limitée.

### 3 Vision prédictive opérationnelle du modèle

L'ensemble des graphiques de la figure 4 présente une méthode graphique de vérification du modèle vers les fortes valeurs :

- La courbe en traits pleins est la " valeur moyenne prédictive " des probabilités de non dépassement des pluies mensuelles d'une année future. Cette valeur moyenne intègre complètement les incertitudes du modèle des fuites.
- Les croix représentent cette valeur moyenne des probabilités en ne faisant que l'hypothèse dite " non paramétrique " sur la distribution des pluies c'est à dire en supposant une distribution quelconque. (Fortin et al, 1997). Le support de ces probabilités est certes la gamme des pluies observées mais celles ci ne sont là que pour la visualisation graphique de valeurs futures possibles. Les courbes doivent donc être interprétées en " prédictif " et non en " retrospectif " comme en statistique classique.

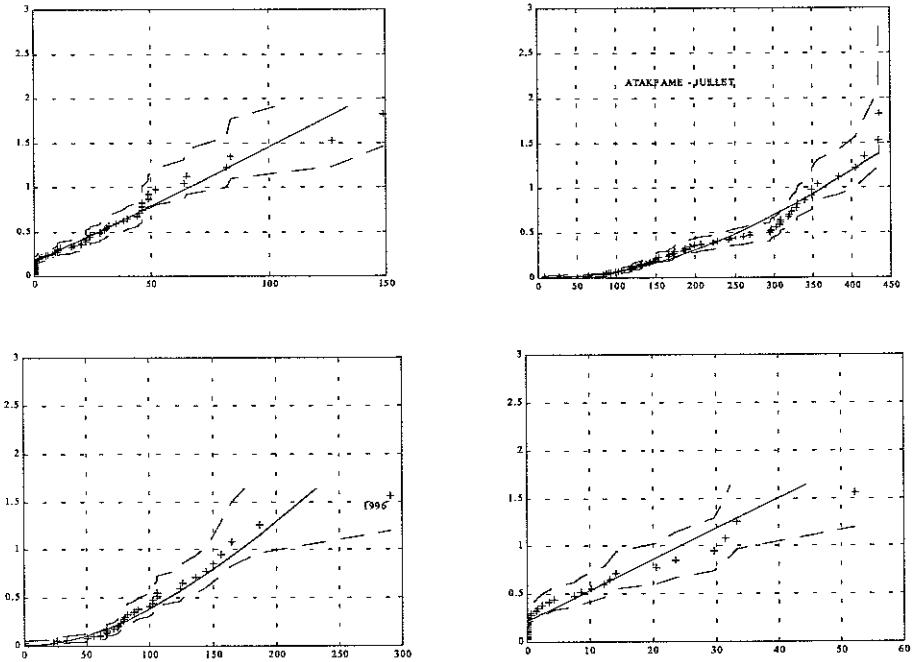


FIG. 4 – Validations prédictives de la loi des fuites

- Cette valeur moyenne est assortie d’une ” fourchette ” courbes inférieure et supérieure en tiretés caractérisant les intervalles de crédibilités à 80% non paramétrique associés à la moyenne prédictive non paramétrique.

Cette validation graphique permet notamment de s’assurer que ce modèle semble particulièrement réaliste pour les mois où la fréquence des pluies nulles est notable. Dans nos deux sites il semble que les autres mois (Juillet pour Atakpame et Février pour Ghezala) des écarts peuvent apparaître pour les fortes précipitations notamment pour la station tunisienne où l’écart est toutefois le fait d’une seule valeur extrême. Ces écarts pourraient être dûs à des erreurs de mesure que nous ignorons ici.

## 4 Conclusions

Les résultats de l'analyse bayesienne a posteriori peuvent être étendus à la vision prédictive du modèle, c'est à dire aux divers aspects de son application à la prévision des précipitations futures y compris à la gestion des ressources qui en découlent. Dans les zones semi-arides, le gestionnaire doit notamment prévoir les possibilités de bas niveau de la ressource. Le modèle des fuites peut ici être utilisé pour "prendre les paris" sur la précipitation mensuelle d'une année future quelconque, ou sur le minimum du mois concerné sur une période de 10 ou de 50 années.

Comme pour tout modèle, dans les applications précédentes à des totaux mensuels, on peut constater certains écarts possibles au modèle des fuites et avoir envie de créer d'autres structures explicatives. L'approche bayesienne complète qui s'appuie sur le "duo gagnant" -modélisation et calculs conditionnels- est assez souple pour s'adapter à de nombreuses généralisations possibles du modèle de renouvellement-dépassement. Par exemple les modèles originaux et intéressants exposés par Cox et Isham (1994) pour la modélisation de la stochasticité des pluies peuvent également faire l'objet de cette approche.

## 5 Bibliographie

**Bernier, J. , 1970.** Inventaire des modèles de processus stochastiques applicables à la description des débits journaliers des rivières. Revue de l'Institut International de Statistique, vol. 38 (1) : 49-61.

**Bernier, J. et D. Fandeux , 1970.** Théorie du renouvellement - Application à l'étude statistique des précipitations mensuelles. Revue de Statistique appliquée, vol. XVIII (2) : 75-87.

**Bernier, J., Parent, E., Boreux, J.J., 2000** : Statistique pour l'environnement. Traitement bayésien de l'incertitude. Lavoisier, Tec & Doc.

**Coles, S. G. , 2001.** *An introduction to statistical modeling of extreme values*. Springer Verlag, New York.

**Cox, D.R. and Isham, V. , 1994.** Stochastic models of precipitation. In Statistics for the environment 2 : Water related issues, (ed. V. Barnett and K. Turkman), pp. 3-18. Wiley, Chichester.

**Fortin, V., Bernier, J. et Bobée, B. , 1997.** Simulation, Bayes, and bootstrap in statistical hydrology. Water Resour. Res., 33 (3) : 439-448.

- Hyfran, 2000** Hydrological Frequency Analysis Software package. Chaire d'hydrologie statistique INRS Eau, programme ([http://www.inrs-eau.quebec.ca/activites/groupes/chaire\\_hydrol](http://www.inrs-eau.quebec.ca/activites/groupes/chaire_hydrol))
- Kuczera, G. and Parent E., 1998.** Monte Carlo assessment of parameter uncertainty in conceptual catchment models : the Metropolis algorithm, *Journal of Hydrology*, 211, 69-85.
- Maidment D.R., 1993** *Handbook of Hydrology*. McGraw-Hill.
- Parent, E. , Bernier, J., 2002.** Bayesian P.O.T. modeling for historical data *Journal of Hydrology*, 274 : 95-108.
- Rivot, E., Prévost, E., Parent, E. , 2001.** How robust are Bayesian inferences based on the Ricker model with regards to measurement errors and to prior assumptions about parameters ?". *Canadian Journal of Fisheries and Aquatic Sciences*. 58 : 2284-2297.
- Robert, C.P. and Casella, G., 1998.** *Monte Carlo statistical methods*, Springer Verlag.
- Tanner, M.A., 1996.** *Tools for statistical inference : methods for the exploration of posterior distribution and likelihood functions*. Springer Verlag, New York.
- Tapsoba D., 1997** Caractérisation événementielle des régimes pluviométriques ouest africains et de leur récent changement. Thèse de doctorat de l'Université Paris-Sud (Orsay)
- Wang, Q.J., 1991.** The POT model described by the generalized Pareto distribution with Poisson arrival rate. *Journal of Hydrology*. 129, 263-280.