

## Article

---

« Comparaison des méthodes d'estimation des paramètres du modèle GEV non stationnaire »

Salaheddine El Adlouni et Taha B.M.J. Ouarda

*Revue des sciences de l'eau / Journal of Water Science*, vol. 21, n° 1, 2008, p. 35-50.

Pour citer cet article, utiliser l'information suivante :

URI: <http://id.erudit.org/iderudit/017929ar>

DOI: 10.7202/017929ar

Note : les règles d'écriture des références bibliographiques peuvent varier selon les différents domaines du savoir.

---

Ce document est protégé par la loi sur le droit d'auteur. L'utilisation des services d'Érudit (y compris la reproduction) est assujettie à sa politique d'utilisation que vous pouvez consulter à l'URI <https://apropos.erudit.org/fr/usagers/politique-dutilisation/>

---

Érudit est un consortium interuniversitaire sans but lucratif composé de l'Université de Montréal, l'Université Laval et l'Université du Québec à Montréal. Il a pour mission la promotion et la valorisation de la recherche. Érudit offre des services d'édition numérique de documents scientifiques depuis 1998.

Pour communiquer avec les responsables d'Érudit : [info@erudit.org](mailto:info@erudit.org)

# COMPARAISON DES MÉTHODES D'ESTIMATION DES PARAMÈTRES DU MODÈLE GEV NON STATIONNAIRE

*Comparison of methods for estimating the parameters of the non-stationary GEV model*

---

SALAHEDDINE EL ADLOUNI et TAHA B.M.J. OUARDA

Chaire en Hydrologie Statistique Hydro-Québec/CRSNG  
Chaire du Canada en estimation des variables hydrologiques  
Institut national de la recherche scientifique, Centre Eau, Terre et Environnement, 490, rue de la Couronne,  
Québec (Québec) G1K 9A9

Reçu le 15 novembre 2005, accepté le 15 juin 2007

---

## RÉSUMÉ

L'analyse fréquentielle des événements extrêmes est un des outils privilégiés pour l'estimation des débits de crue et de leurs périodes de retour. En analyse fréquentielle, les observations doivent être indépendantes et identiquement distribuées (iid). Ces hypothèses ne sont pas souvent respectées et les paramètres de la loi à ajuster sont fonction du temps ou de covariables. Le modèle GEV non stationnaire permet de tenir compte de cette dépendance. L'objectif du présent travail est de comparer la méthode du maximum de vraisemblance pour l'estimation des quantiles à la méthode du maximum de vraisemblance généralisée (GML) et à une généralisation de la méthode des L-moments dans le cas non stationnaire. Trois modèles sont considérés : le modèle stationnaire (GEV0), le cas où le paramètre de position est une fonction linéaire de la covariable (GEV1) et le cas d'une dépendance quadratique (GEV2). Un cas d'étude des précipitations à une station de la Californie montre le potentiel des modèles non stationnaires.

**Mots-clés :** Valeurs extrêmes, maximum de vraisemblance généralisé, L-moments, Indice d'oscillations du sud.

## ABSTRACT

In frequency analysis, data must generally be independent and identically distributed (i.i.d), which implies that they must meet the statistical criteria of independence, stationarity and homogeneity. In reality, the probability distribution of extreme events can change with time, indicating the existence of non-stationarity. The objective of the present study was to develop efficient estimation methods for the use of the GEV distribution for quantile estimation in the presence of non-stationarity. Parameter estimation in the non-stationary GEV model is generally done with the Maximum Likelihood Estimation method. In this work, we suggest two other estimation methods: the Generalized Maximum Likelihood Estimation (GML) and the generalization of the L-moment method for the non-stationary case. A simulation study was carried out to compare the performances of these three estimation methods in the case of the stationary GEV model (GEV0), the non-stationary case with a linear dependence (GEV1), and the non-stationary case with a quadratic dependence on covariates (GEV2). The non-stationary GEV model was also applied to a case study from the State of California to illustrate its potential.

**Keywords :** Extreme value, Generalized maximum likelihood, L-moments, Southern oscillation index.

---

\*Auteur pour correspondance :

Téléphone: 418 654-2639

Télécopie: 418 654-2600 Courriel :

Courriel: salaheddine\_eladlouni@ete.inrs.ca

## 1. INTRODUCTION

Les valeurs extrêmes sont souvent les maxima d'une certaine quantité sur une période donnée. La théorie des valeurs extrêmes montre que la distribution de ces maxima est représentée par une des trois lois des valeurs extrêmes. La loi généralisée des valeurs extrêmes (GEV) introduite par JENKINSON (1955) est une distribution à trois paramètres qui combine les trois types en une seule forme. En hydrologie, la loi GEV est parmi les distributions les plus utilisées en analyse des extrêmes (KATZ *et al.*, 2002; STEDINGER *et al.*, 1993). Les méthodes les plus connues pour l'estimation des paramètres de la loi GEV sont : la méthode du maximum de vraisemblance (MV), la méthode des moments (MM) et la méthode des L-moments (LM) qui est équivalente à la méthode des moments de probabilités pondérées (PWM). Pour résoudre le problème de divergence des méthodes numériques utilisées dans la méthode MV, MARTINS et STEDINGER (2000) ont suggéré l'emploi d'une loi *a priori* pour restreindre le domaine d'appartenance du paramètre de forme de la loi GEV à un intervalle de valeurs les plus probables. Cette approche est équivalente à la méthode QB-MLE (Quasi-Bayesian Maximum Likelihood Estimator), proposée par HAMILTON (1991). Cette méthode a été appliquée par VENKATARAMAN (1997) en finance. La méthode QB-MLE est une extension de la méthode MV, où on tient compte de l'information *a priori* sur les paramètres pour éliminer les singularités associées à la méthode MV. MORRISON et SMITH (2003) ont introduit une nouvelle méthode qu'ils ont appelé « mixed L-Moments : Maximum Likelihood » pour résoudre le même problème et pour obtenir des estimateurs sans biais comme ceux donnés par la méthode MV. Cette méthode consiste à résoudre les équations liées à la méthode du maximum de vraisemblance sous des contraintes données par le premier L-moment ou les deux premiers L-moments. Ils montrent, par simulation, que les estimateurs obtenus par cette méthode conservent toujours la propriété des estimateurs sans biais et possèdent une faible variance.

En analyse fréquentielle, en plus des problèmes liés à l'estimation des paramètres, la série de données doit vérifier certaines hypothèses comme l'indépendance, la stationnarité et l'homogénéité. Ces hypothèses, qui sont équivalentes à la notion d'observations indépendantes et identiquement distribuées (i.i.d.), sont à la base de toute analyse fréquentielle. En réalité, la loi de probabilité des crues peut changer avec le temps (existence de non-stationnarité), conséquence des activités humaines ou des changements climatiques. Le rapport du « Intergovernmental Panel for Climate Change » (IPCC 2001) conclut à une augmentation de la température du globe avec des effets sur la fréquence des événements extrêmes. Ce changement, soit dans la tendance (ZHANG *et al.*, 2001) ou au niveau de l'échelle des valeurs extrêmes, met en question l'hypothèse de stationnarité des séries hydrologiques qui est une hypothèse indispensable dans l'analyse fréquentielle

(OUARDA *et al.*, 1999). Plusieurs tests permettent de vérifier cette hypothèse pour une série d'observations donnée. Les plus utilisés sont le test de Mann-Kendall et Spearman (CHEN et RAO, 2002; ÖNÖZ et BAYAZIT 2003; YUE *et al.*, 2002).

L'effet des changements climatiques sur les variables hydrologiques liées à des événements extrêmes (pluie maximale annuelle, débit maximum annuel, etc.) peut se faire par l'étude de l'existence de tendances dans les séries observées, ou en analysant la dépendance des variables étudiées avec d'autres variables ou indices climatiques, appelées covariables (KATZ, 1999). CLARKE (2002) a présenté la méthode du maximum de vraisemblance pour l'estimation des paramètres de la loi Weibull dans le cas d'une tendance. SANKARASUBRAMANIAN et LALL (2003) ont étudié la régression des quantiles avec des indices climatiques pour l'estimation des quantiles dans le cas des changements climatiques. Le modèle GEV non stationnaire a été introduit dans le cas d'une tendance au niveau du paramètre de position par SCARF (1992), qui a présenté la méthode du maximum de vraisemblance pour l'estimation des paramètres du modèle GEV à quatre paramètres. COLES (2001) a présenté le modèle GEV non stationnaire sous une forme plus générale et la méthode du maximum de vraisemblance pour l'estimation des paramètres. Le modèle GEV non stationnaire, tel qu'il est présenté dans COLES (2001) est général. KHARIN et ZWIERS (2005) et WANG *et al.* (2004) ont appliqué le même modèle sur des données climatiques simulées.

L'estimation des paramètres du modèle GEV non stationnaire est généralement faite par la méthode du maximum de vraisemblance. L'approche bayésienne permet d'intégrer des informations additionnelles dans le processus d'estimation, ce qui permet de combler le manque d'information dû aux courtes périodes d'observation ou tout simplement de résoudre les problèmes de calcul numérique. Elle constitue donc une bonne alternative à la méthode du maximum de vraisemblance quand on a une information *a priori* sur les paramètres. Elle évite également les problèmes numériques rencontrés dans la résolution du système d'équations liées à cette dernière méthode, grâce à l'évolution des méthodes de Monte Carlo par Chaînes de Markov (MCMC). Basée sur le même principe que l'approche bayésienne, la méthode du maximum de vraisemblance généralisé, introduite par MARTINS et STEDINGER (2000), considère une loi *a priori* sur le paramètre de forme de la loi GEV. Cette même approche a été étendue au cas non stationnaire par EL ADLOUNI *et al.* (2007). Ils l'ont comparée à la méthode du maximum de vraisemblance dans le cas d'une non-stationnarité au niveau de la moyenne.

L'objectif de ce travail est de présenter une généralisation de la méthode des L-moments à certains cas du modèle GEV non stationnaire. Cette approche sera comparée à la méthode du maximum de vraisemblance et la méthode du maximum de

vraisemblance généralisé. La comparaison est basée sur le biais et la racine de l'erreur quadratique moyenne des quantiles.

Dans la section suivante, on présentera le modèle GEV non stationnaire sous une forme générale. On présentera ensuite un rappel sur les méthodes d'estimation des paramètres dans le cas non stationnaire et la généralisation de l'approche des L-moments. La section 4 est consacrée à la comparaison des méthodes d'estimation par simulations de Monte Carlo. Une illustration de la flexibilité de ces modèles non stationnaires pour l'estimation des quantiles conditionnels de la pluie maximale annuelle est présentée à la section 5. La conclusion et la discussion générale sont présentées à la section 6.

## 2. LE MODÈLE GEV NON STATIONNAIRE

Les distributions des valeurs extrêmes, introduites par FISHER et TIPPETT (1928), comprennent trois lois. Le terme « valeurs extrêmes » est associé à ces distributions parce qu'elles peuvent être obtenues comme lois limites (quand la taille de l'échantillon est grande) de la valeur maximale de  $m$  variables aléatoires indépendantes et identiquement distribuées. JENKINSON (1955) a proposé ce qu'il a appelé la loi Généralisée des Valeurs Extrêmes (GEV). Cette distribution a l'avantage de combiner les trois types de lois des valeurs extrêmes développées par FISHER et TIPPETT (1928). Pour une variable aléatoire, distribuée suivant une loi GEV, la fonction de distribution s'écrit :

$$F_{\text{GEV}}(x) = \exp\left[-\left(1 - \frac{\kappa}{\alpha}(x - \mu)\right)^{1/\kappa}\right] \quad \kappa \neq 0$$

$$= \exp\left[-\exp\left(-\frac{(x - \mu)}{\alpha}\right)\right] \quad \kappa = 0$$
(1)

où  $\mu + \alpha / \kappa \leq x < \infty$  pour  $\kappa < 0$ ,  $-\infty < x < +\infty$  pour  $\kappa = 0$  et  $-\infty < x \leq \mu + \alpha / \kappa$  pour  $\kappa > 0$ . Avec  $\mu (\in \mathbb{R})$ ,  $\alpha (> 0)$  et  $\kappa (\in \mathbb{R})$  sont respectivement les paramètres de position, d'échelle et de forme.

Dans le cas de la non-stationnarité, les paramètres ne sont pas des constantes mais dépendent du temps ou d'autres covariables ( $\text{GEV}(\mu_t, \alpha_t, \kappa_t)$ ). Pour éviter les problèmes de valeurs nulles du paramètre d'échelle  $\alpha_t$ , on considère la paramétrisation  $\varphi_t = \log(\alpha_t)$ . On se restreint aux cas de dépendances linéaires. Cependant, plusieurs autres types de dépendances peuvent être ramenés au cas linéaire.

Soit  $U = (U_1 \ U_2 \ \dots \ U_{n_\mu})'$  le vecteur de  $n_\mu$  covariables du paramètre de position  $\mu_t$  et :

$$\mu_t = U'(t)\beta = \sum_{i=1}^{n_\mu} \beta_i U_i(t) \quad (2)$$

où  $\beta = (\beta_1 \ \beta_2 \ \dots \ \beta_{n_\mu})'$  est le vecteur des paramètres du modèle de dépendance du paramètre  $\mu_t$  du vecteur de covariables  $U$ .

Pour le paramètre d'échelle  $\alpha_t$ , soit  $V = (V_1 \ V_2 \ \dots \ V_{n_\alpha})'$ , le vecteur des covariables est :

$$\varphi_t = \log(\alpha_t) = V'(t) \cdot \delta = \sum_{i=1}^{n_\alpha} \delta_i V_i(t) \quad (3)$$

avec  $\delta = (\delta_1 \ \delta_2 \ \dots \ \delta_{n_\alpha})'$ , le vecteur des paramètres.

De la même façon, le paramètre de forme peut s'écrire en fonction d'un vecteur de covariables  $W = (W_1 \ W_2 \ \dots \ W_{n_\kappa})'$  et un vecteur des hyper-paramètres  $\gamma = (\gamma_1 \ \gamma_2 \ \dots \ \gamma_{n_\kappa})'$  :

$$\kappa_t = W'(t) \cdot \gamma = \sum_{i=1}^{n_\kappa} \gamma_i W_i(t) \quad (4)$$

Pour un échantillon observé  $\underline{x} = \{x_1, \dots, x_n\}$  de la variable étudiée  $X$ , la fonction de vraisemblance dans le cas de paramètres dépendants de covariables s'écrit de la façon suivante :

$$L_n = \prod_{t=1}^n f(x_t; \mu_t, \varphi_t, \kappa_t) \quad (5)$$

$f$  est la fonction de densité de probabilité de la loi GEV et correspond à la dérivée de la fonction de répartition donnée à l'équation 1.

En pratique, les modèles les plus intéressants sont ceux qui représentent : (i) la non-stationnarité au niveau de la moyenne avec une variance constante (modèle homoscedastique), et (ii) la non-stationnarité au niveau de la moyenne et la variance (modèle hétéroscedastique). Les modèles étudiés dans le présent travail concernent surtout le paramètre de position et sont :

$\text{GEV}_0(\mu, \alpha, \kappa)$  Modèle classique : tous les paramètres sont constants.  $\mu_t = \mu$ ,  $\alpha_t = \alpha$  et  $\kappa_t = \kappa$ . Dans ce cas,  $n_\mu = n_\alpha = n_\kappa = 1$  et  $U_1 = V_1 = W_1 = 1$ .

$\text{GEV}_1(\mu_t = \beta_1 + \beta_2 Y_t, \alpha, \kappa)$ . Existence d'une tendance dans un modèle homoscedastique : le paramètre de position dépend

d'une variable temporelle ( $Y_t$ ) et les deux autres paramètres sont constants. Dans ce cas,

$$n_\mu = 2, U(t) = (U_1(t) = 1 \quad U_2(t) = Y_t), n_\alpha = n_\kappa = 1 \text{ et } V_1 = W_1 = 1$$

GEV<sub>2</sub> ( $\mu_t = \beta_1 + \beta_2 Y_t + \beta_3 Y_t^2, \alpha, \kappa$ ). Le paramètre de position est une fonction quadratique d'une covariable temporelle  $Y_t$ . Les deux autres paramètres sont constants. Dans ce cas,

$$n_\mu = 3, U(t) = (U_1(t) = 1 \quad U_2(t) = Y_t \quad U_3(t) = Y_t^2),$$

$$n_\alpha = n_\kappa = 1 \text{ et } V_1 = W_1 = 1.$$

### 3. ESTIMATION DES PARAMÈTRES

Soit  $\underline{x} = (x_1, \dots, x_n)$ , le vecteur des observations de la variable  $X$  et  $\theta$  le vecteur des paramètres à estimer. Dans notre cas :  $\theta = (\mu, \alpha, \kappa)$  pour le modèle GEV0, pour le modèle GEV1,  $\theta = (\beta_1, \beta_2, \alpha, \kappa)$  et pour le modèle GEV2,  $\theta = (\beta_1, \beta_2, \beta_3, \alpha, \kappa)$ .

#### 3.1 Méthode du maximum de vraisemblance

Les paramètres du modèle GEV non stationnaire peuvent être estimés par la méthode du maximum de vraisemblance qui s'écrit d'une manière générale :

$$L_n(\underline{x} | \mu_t, \alpha_t, \kappa_t) = \prod_{t=1}^{n_1} \frac{1}{\alpha_t} \exp \left\{ - \left[ 1 - \kappa_t \left( \frac{x_t - \mu_t}{\alpha_t} \right) \right]^{-1/\kappa_t} \right\} \\ * \left[ 1 - \kappa_t \left( \frac{x_t - \mu_t}{\alpha_t} \right) \right]^{-\left(1 - \frac{1}{\kappa_t}\right)} \\ * \prod_{t=n_1+1}^n \frac{1}{\alpha_t} \exp \left\{ - \left( \frac{x_t - \mu_t}{\alpha_t} \right) \right\} * \exp \left\{ - \exp \left[ - \left( \frac{x_t - \mu_t}{\alpha_t} \right) \right] \right\} \quad (6)$$

$n_1$  est le nombre d'observations telles que  $\kappa_t \neq 0$ .

Pour les modèles étudiés dans ce travail, la non-stationnarité est liée au paramètre de position. Dans ce cas, et pour tous les modèles considérés,  $\alpha_t = \alpha$  et  $\kappa_t = \kappa$  sont constants. Lorsque  $\kappa \neq 0$  et  $n_1 = n$ , la fonction log-vraisemblance est :

$$l_n(\underline{x}; \mu_t, \alpha, \kappa) = -n \log(\alpha) - \sum_{t=1}^n \left[ 1 - \kappa \left( \frac{x_t - \mu_t}{\alpha} \right) \right]^{1/\kappa} \\ - \sum_{t=1}^n \left( 1 - \frac{1}{\kappa} \right) \log \left[ 1 - \kappa \left( \frac{x_t - \mu_t}{\alpha} \right) \right] \quad (7)$$

En pratique, il est plus facile de maximiser la fonction log-vraisemblance. Les estimateurs par maximum de vraisemblance sont donc les solutions du système d'équations correspondant aux dérivées partielles de  $l_n$  par rapport à chacun des paramètres.

Pour le modèle GEV1, les estimateurs du maximum de vraisemblance des paramètres  $(\beta_1, \beta_2, \alpha, \kappa)$  sont les solutions du système d'équations :

$$\sum_{i=1}^n \frac{1 - \kappa - z_t^{1/\kappa}}{z_t} = 0 \\ \sum_{i=1}^n t \frac{1 - \kappa - z_t^{1/\kappa}}{z_t} = 0 \\ -n + \sum_{i=1}^n \left[ \frac{1 - \kappa - z_t^{1/\kappa}}{z_t} \left( \frac{x_t - \mu_t}{\alpha} \right) \right] = 0 \\ \sum_{i=1}^n \left\{ \ln(z_t) \left[ 1 - \kappa - z_t^{1/\kappa} \right] + \frac{1 - \kappa - z_t^{1/\kappa}}{z_t} \kappa \left( \frac{x_t - \mu_t}{\alpha} \right) \right\} = 0 \quad (8)$$

où  $z_t = \left[ 1 - \frac{\kappa}{\alpha} (x_t - \mu_t) \right]$ . La résolution de ce système peut se faire d'une manière numérique en utilisant la méthode de Newton-Raphson. De la même façon, on obtient un système équivalent à (8) pour le modèle GEV2, avec une cinquième équation correspondant au paramètre  $\beta_3$ .

Les estimateurs du maximum de vraisemblance, sous certaines conditions de régularités sont asymptotiquement efficaces. Ces conditions de régularité ne sont pas vérifiées lorsque le paramètre de forme est non nul, puisque le support de la distribution dépend des paramètres (SMITH, 1985). Des approximations de l'erreur standard et des intervalles de confiance des estimateurs peuvent être obtenus comme dans le cas classique par une évaluation numérique de la matrice d'information observée. Pour les échantillons de petites tailles, la résolution numérique du système d'équations (8) peut conduire à des estimateurs des paramètres physiquement impossibles et cause un biais et une variance élevée des quantiles extrêmes. Plusieurs alternatives ont été proposées dans la littérature pour résoudre ce problème, en particulier la méthode des

L-moments (HOSKING, 1990) et la méthode du maximum de vraisemblance généralisée (MARTINS et STEDINGER, 2000). On présente ci-dessous des généralisations de ces deux méthodes aux cas non stationnaires.

### 3.2 Méthode des L-Moments

La méthode des L-moments (LM), introduite par HOSKING (1990), est équivalente à la méthode des moments de probabilité pondérés (PWM). Les estimateurs par L-moments des paramètres du modèle GEV classique, lorsque  $-0,5 < \kappa < 0,5$ , sont donnés par :

$$\hat{k} \approx 7,8590c + 2,9554c^2 \quad (9)$$

$$\hat{\alpha} = \frac{l_2 \hat{k}}{(1-2^{-\hat{k}}) \Gamma(1+\hat{k})} \quad (10)$$

$$\hat{\mu} = l_1 - \hat{\alpha} \left\{ 1 - \Gamma(1+\hat{k}) \right\} / \hat{k} \quad (11)$$

où  $c = \frac{2}{3+t_3} - \frac{\log 2}{\log 3}$  et  $l_1, l_2$  et  $l_3$  sont, respectivement, les estimateurs des L-moments d'ordres 1 et 2 et du rapport des L-moments d'ordre 3. Ces estimateurs peuvent être définis à partir des moments de probabilité pondérés (GREENWOOD *et al.*, 1979) :

$$\beta_r = E \left\{ X [F(x)]^r \right\} \quad (12)$$

Un estimateur sans biais de  $\beta_r$  ( $r \geq 0$ ) est :

$$b_0 = n^{-1} \sum_{j=1}^n X_j \quad (13)$$

$$b_r = n^{-1} \sum_{j=r+1}^n \frac{(j-1)(j-2)\dots(j-r)}{(n-1)(n-2)\dots(n-r)} X_j \quad (14)$$

et les estimateurs des quatre premiers L-moments sont :

$$l_1 = b_0 \quad (15)$$

$$l_2 = 2b_1 - b_0 \quad (16)$$

$$l_3 = 6b_2 - 6b_1 + b_0 \quad (17)$$

$$l_4 = 20b_3 - 30b_2 + 12b_1 - b_0 \quad (18)$$

En divisant les L-moments d'ordre supérieur à 2 par la mesure de dispersion  $l_2$ , on obtient les estimateurs des rapports des L-moments  $t_r = \frac{l_r}{l_2}$   $r = 3, 4, \dots$

En pratique, le coefficient de la L-asymétrie ( $t_3$ ) peut s'écrire d'une manière directe de la façon suivante :

$$t_3 = \sum_{i=1}^n \frac{c_i X_{(i)} + \bar{X}}{l_2} \quad (19)$$

où  $c_i = 6 \frac{(i-1)(i-2)}{n(n-1)(n-2)} - 6 \frac{(i-1)}{n(n-1)}$  et  $\bar{X}$  étant la moyenne de l'échantillon.

Cette approche peut être généralisée aux cas non stationnaires lorsque seul le paramètre de position est fonction de covariables. En effet, si la variable étudiée  $X$  est distribuée suivant une loi GEV1( $\mu_t, \alpha, \kappa$ ) alors

$$\begin{aligned} E[X] &= \mu_t + \frac{\alpha}{\kappa} \{1 - \Gamma(1 + \kappa)\} \\ &= \beta_1 + \beta_2 Y_t + \frac{\alpha}{\kappa} \{1 - \Gamma(1 + \kappa)\} \end{aligned} \quad (20)$$

Un estimateur  $\beta_2$  de  $\beta_2$  peut être obtenu à partir d'une régression linéaire simple de la variable  $X$  par rapport à la covariable  $Y_t$ . Ensuite, avec un changement de variable  $S = X - \beta_2 Y_t$ , on obtient une variable distribuée suivant une loi GEV( $\beta_1, \alpha, \kappa$ ). En effet, étant donné que le paramètre  $\mu_t$  est un paramètre de position, la distribution GEV est stable par rapport à toute translation. Les paramètres  $\beta_1, \alpha$  et  $\kappa$  de la nouvelle variable sont ensuite estimés par la méthode des L-moments à partir des équations 9, 10 et 11 appliquées à l'échantillon transformé correspondant à la variable  $S$ . La même approche peut être utilisée dans le cas du modèle quadratique où les deux paramètres,  $\beta_2$  et  $\beta_3$ , liés à la dépendance linéaire et quadratique, sont estimés par une régression linéaire multiple par rapport à  $Y_t$  et  $Y_t^2$ .

### 3.3 Méthode du maximum de vraisemblance généralisée

La méthode du maximum de vraisemblance généralisée (GML) est basée sur le même principe que la méthode du maximum de vraisemblance avec une contrainte additionnelle sur le paramètre de forme pour restreindre le domaine des solutions à des valeurs qui sont physiquement acceptables. Une loi *a priori* pour le paramètre de forme  $k$ , dans le cas des séries

hydrologiques, a été introduite par MARTINS et STEDINGER (2000) avec des considérations pratiques. L'objectif est d'éliminer les problèmes de solutions numériques qui n'ont pas de signification physique. MARTINS et STEDINGER (2000) ont présenté la méthode GML au cas du modèle GEV0 avec une loi *a priori* pour le paramètre de forme qui est une loi Beta de paramètre  $u = 6$  et  $v = 9$  :  $\pi_{\kappa}(\kappa) = \text{Beta}(u = 6, v = 9)$ . Il s'agit d'une loi *a priori* centrée sur la valeur  $-0,1$  et de support  $[-0,5, +0,5]$ . Cette méthode a été généralisée au cas non stationnaire par EL ADLOUNI *et al.* (2007) en considérant la même loi *a priori* pour le paramètre de forme. Le système d'équations du maximum de vraisemblance (équation (8) pour le modèle GEV1) peut ensuite être résolu sous cette contrainte sur le paramètre de forme. Les estimateurs GML des paramètres sont donc des solutions du problème d'optimisation suivant :

$$\begin{cases} \max_{\theta} L_n(\underline{x}; \theta) \\ \kappa \sim \text{Beta}(u, v) \end{cases} \quad (21)$$

Ce qui revient à maximiser la loi *a posteriori* des paramètres conditionnellement aux données :

$$\pi(\theta | \underline{x}) \propto L_n(\underline{x} | \theta) \pi_{\kappa}(\kappa) \quad (22)$$

où  $L_n$  est donné par l'équation 6 et  $\pi_{\kappa}$  est la loi *a priori* du paramètre de forme  $\kappa$ .

Comme il a été signalé dans MARTINS et STEDINGER (2000), la méthode GML est un cas particulier de l'approche bayésienne où la seule loi *a priori* considérée concerne le paramètre de forme. L'estimateur GML des paramètres correspond au mode de la distribution *a posteriori* :

$$\pi(\theta | \underline{x}) = \frac{L_n(\underline{x} | \theta) \pi(\theta)}{\int L_n(\underline{x} | \theta) \pi(\theta) d\theta} \quad (23)$$

qui n'est pas connu de manière explicite. Le calcul des estimateurs peut se faire par des méthodes numériques (MARTINS et STEDINGER, 2000) ou par des méthodes de simulation comme les méthodes de Monte Carlo par Chaînes de Markov (MCMC). Les méthodes MCMC permettent d'obtenir la distribution *a posteriori* empirique du vecteur des paramètres et d'en déduire les lois marginales des paramètres et leurs caractéristiques. Les estimateurs GML sont donc les modes des lois marginales empiriques obtenues par les méthodes MCMC.

## 4. TECHNIQUES MCMC POUR LE CALCUL DES ESTIMATEURS ET DES INTERVALLES DE CRÉDIBILITÉ

Les méthodes de Monte Carlo par Chaînes de Markov (MCMC) constituent une alternative aux méthodes numériques pour la résolution du système d'équations obtenues par la méthode GML. Les méthodes MCMC sont souvent utilisées pour résoudre les problèmes de calcul liés à l'approche bayésienne. Ce sont des méthodes de simulation puissantes qui permettent de générer des observations issues de la loi *a posteriori* explicite à une constante de normalisation près. L'algorithme MCMC permet de générer une chaîne de Markov de taille  $N$ , ayant comme loi stationnaire la loi d'intérêt. Après un certain nombre d'itérations,  $N_0$ , qu'on appelle temps de chauffe, les états générés sont des réalisations de la loi stationnaire qui, dans le cas de l'estimation bayésienne, et en particulier GML, correspond à la loi *a posteriori*  $\pi(\theta | \underline{x})$ . Le choix du temps de chauffe et de la longueur de la chaîne générée constitue une étape importante dans le processus d'estimation pour assurer la convergence des chaînes de Markov vers la loi stationnaire. Plusieurs méthodes ont été développées pour déterminer le temps de chauffe et la longueur de la chaîne ou pour vérifier la convergence. Une étude détaillée sur ces méthodes est donnée par EL ADLOUNI *et al.* (2006).

Les estimateurs des paramètres du modèle GEV non stationnaire par la méthode GML sont obtenus par simulation de la distribution *a posteriori* à partir de l'algorithme de Metropolis-Hastings (M-H) (HASTINGS, 1970; METROPOLIS *et al.*, 1953). Chaque paramètre est mis à jour en utilisant la version « marche aléatoire » de l'algorithme M-H avec une distribution instrumentale gaussienne centrée sur l'état courant de la chaîne. Dans tous les cas considérés dans l'étude par simulation, le graphique de la chaîne obtenue à partir de l'algorithme M-H montre que la chaîne converge vers la distribution stationnaire après peu d'itérations. Nous avons considéré des chaînes de taille  $N = 10\,000$  avec un temps de chauffe  $N_0 = 4\,000$ . L'estimateur GML correspond au mode de la distribution *a posteriori*, tandis que l'estimateur de Bayes correspond à la moyenne, obtenus à partir des  $N - N_0$  valeurs générées après le temps de chauffe  $N_0$ .

En plus des estimateurs des paramètres, les itérations de l'algorithme M-H permettent d'obtenir la distribution conditionnelle des quantiles par rapport à une valeur observée de la covariable  $Y_t$ . En effet, pour chaque itération de l'algorithme M-H,  $i = 1, \dots, N$ , on calcule le quantile de probabilité de non-dépassement  $p$ ,  $x_{p, y_0}^{(i)}$  correspondant au vecteur des paramètres  $(\mu_{y_0}^{(i)}, \alpha^{(i)}, \kappa^{(i)})$ , à l'aide de l'inverse de la fonction de répartition de la loi GEV. On en déduit le quantile  $x_{p, y_0}^{(i)}$  conditionnellement à la valeur  $y_0$  de  $y_t$  :

$$x_{p,y_0}^{(i)} = \left( \mu_{y_0}^{(i)} \right) + \frac{\alpha^{(i)}}{\kappa^{(i)}} \left[ 1 - (-\log(p))^{k^{(i)}} \right] \quad (24)$$

où  $\mu_{y_0}^{(i)}$  est le paramètre de position conditionnellement à une valeur particulière  $y_0$  de  $y_t$  :  $\mu_{y_0}^{(i)} = \mu^{(i)}$  pour le modèle GEV0, pour le modèle GEV1 et dans le cas du modèle GEV2  $\mu_{y_0}^{(i)} = \beta_1^{(i)} + \beta_2^{(i)} y_0 + \beta_3^{(i)} y_0^2$ .

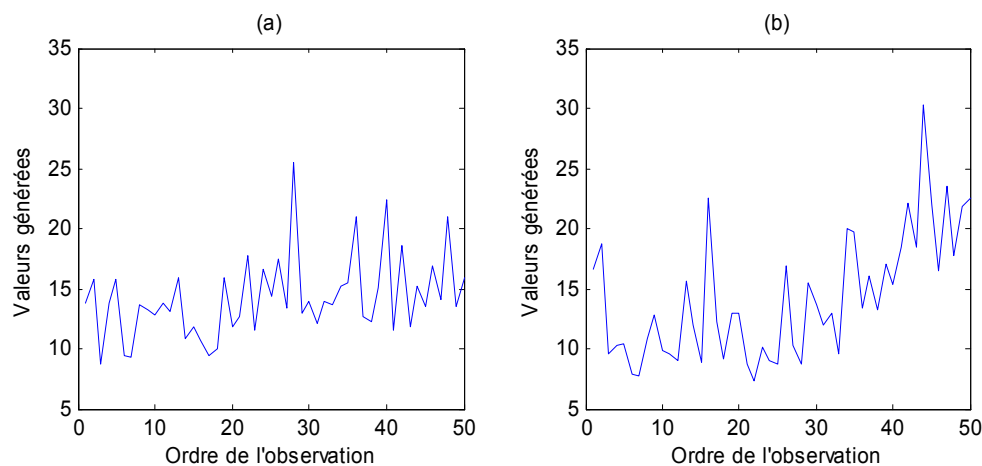
Plusieurs caractéristiques de la distribution conditionnelle de chaque quantile peuvent être déterminées à partir des valeurs  $x_{p,y_0}^{(i)}$ ,  $i = N_0, \dots, N$ , comme la moyenne, le mode ou un intervalle de confiance, etc.

## 5. COMPARAISON DES MÉTHODES PAR SIMULATION

MARTINS et STEDINGER (2000) ont comparé la méthode MV à d'autres méthodes d'estimation : la méthode LM, la méthode des moments (MOM) et la méthode GML pour le modèle GEV stationnaire (GEV0). Ils ont constaté que la méthode GML donne les meilleurs résultats au niveau du biais et de la racine de l'erreur quadratique moyenne des quantiles même dans le cas des échantillons de petite taille. Dans cette section, on compare les trois méthodes d'estimation des paramètres (MV, GML et LM) présentées dans ce travail, pour les modèles non stationnaires GEV1 et GEV2. La

covariable considérée, dans cette étude par simulation, est le temps  $y_t = t$ . Le quantile de probabilité de non-dépassement  $p$  est calculé pour  $Y_t = n$  ( $x_{p,n}$ ), ce qui correspond à une mise à jour de la distribution de la variable étudiée, par rapport à la dernière valeur de la covariable  $Y_t$  (temps). Nous avons repris les paramètres de la loi GEV considérés dans MARTINS et STEDINGER (2000) pour comparer la méthode GML à la méthode des L-moments et la méthode du maximum de vraisemblance. Nous nous sommes intéressés aux asymétries positives avec les valeurs du paramètre de forme suivantes :  $\kappa = -0,1$ ,  $\kappa = -0,2$  et  $\kappa = -0,3$ . L'objectif est de comparer les différentes méthodes d'estimation en fonction des valeurs du paramètre de forme et pour une taille d'échantillon  $n = 50$ . Les paramètres représentant la non-stationnarité sont choisis pour avoir une faible tendance, ce qui correspond à la majorité des séries rencontrées en hydrologie. Pour le modèle GEV1 :  $\beta_1 = 10$  et  $\beta_2 = 0,1$ , et pour le modèle GEV2 :  $\beta_1 = 10$ ,  $\beta_2 = -0,1$  et  $\beta_3 = 0,005$ . Le paramètre d'échelle est tel que  $\alpha = 1$  pour tous les modèles. La figure 1 illustre les séries générées à partir des modèles GEV1 et GEV2.

La comparaison est basée sur le biais d'estimation et la racine de l'erreur quadratique moyenne (REQM) des quantiles de probabilité de non-dépassement  $p = 0,5, 0,8, 0,9, 0,99$  et  $0,999$ , ce qui correspond dans le cas stationnaire aux périodes de retour 2, 5, 10, 100 et 1 000 ans. On préfère parler des quantiles au lieu de la notion de période de retour qui est difficile à interpréter dans le cas non stationnaire. Le biais et le REQM sont obtenus pour  $R = 1\ 000$  échantillons de taille  $n = 50$  générés à partir de chacun des modèles étudiés.



**Figure 1. Séries générées à partir des modèles (a) GEV1 ( $\beta_1 = 10, \beta_2 = 0,1, \alpha = 1, \kappa = -0,1$ ) et (b) GEV2 ( $\beta_1 = 10, \beta_2 = -0,1, \beta_3 = 0,005, \alpha = 1, \kappa = -0,1$ ).**  
*Generated data from the GEV1 and GEV2 models.*



**Modèle GEV1**

Le modèle GEV1 représente le cas où le paramètre de position est une fonction linéaire d'une variable temporelle (covariable). Pour l'étude par simulation de ce modèle, la covariable  $Y_t$  représente le temps et nous avons considéré les paramètres mentionnés ci-dessus, soit  $\beta_1 = 10$ ,  $\alpha = 1$  et le paramètre de la composante linéaire  $\beta_2 = 0,1$ . Trois cas d'asymétries positives ( $\kappa = -0,1$ ,  $\kappa = -0,2$  et  $\kappa = -0,3$ ) en été considérés.

Le tableau 1, montre que la méthode GML donne les meilleurs résultats pour  $k = -0,1$ . La différence entre les trois méthodes est plus significative pour les quantiles extrêmes. Les méthodes GML et LM donnent des résultats similaires au niveau du biais absolu. On remarque que pour les fortes asymétries ( $k \leq -0,2$ ) et dans le cas des quantiles extrêmes ( $p \geq 0,99$ ), les estimateurs GML ont un biais négatif élevé par rapport aux deux autres méthodes. Ceci est dû au fait que cette méthode considère une loi *a priori* centrée en  $-0,1$  pour

le paramètre de forme. Cependant, le REQM est faible par rapport aux méthodes MV et LM.

**Modèle GEV2**

Le modèle GEV2 représente le cas où le paramètre de position est une fonction quadratique d'une covariable (temps dans ce cas). Ce dernier cas montre la flexibilité du modèle GEV non stationnaire pour étudier d'autres types de dépendance. Les résultats obtenus pour le modèle GEV2, pour les trois méthodes, sont présentés dans le tableau 2.

On remarque que pour la méthode GML, le fait de considérer une dépendance quadratique a donné moins de poids à la valeur centrale de la loi *a priori* du paramètre  $k$ . Le biais correspondant est le plus faible pour les trois valeurs en comparaison aux méthodes MV et LM. On remarque aussi que le biais des estimateurs du maximum de vraisemblance est négatif pour la majorité des quantiles et pour ces trois asymétries mais plus grand en valeur absolue, par comparaison aux deux autres méthodes. Cette augmentation du biais est associée à

**Tableau 1. Biais et REQM des quantiles estimés par la méthode MV, LM et GML pour le modèle GEV1.**

**Table 1. Bias and Root mean square error of the quantile estimators given by the Maximum likelihood, L-moment and Generalized maximum likelihood methods for the GEV1 model.**

$\rho$	Bias			RMSE		
	MV	LM	GML	MV	LM	GML
k = -0,1						
0,5	0,1	-0,008	0,02	0,59	0,15	0,10
0,8	0,18	-0,03	0,06	0,65	0,21	0,18
0,9	0,16	-0,06	0,1	0,75	0,30	0,29
0,99	0,30	-0,16	0,18	3,27	2,44	1,87
0,999	0,81	-0,4	0,63	23,9	18,38	11,18
k = -0,2						
0,5	0,34	0,05	0,07	1,05	0,26	0,10
0,8	0,35	0,05	0,05	1,15	0,36	0,16
0,9	0,36	0,04	-0,02	1,38	0,57	0,25
0,99	0,70	0,12	-0,75	8,79	8,20	2,57
0,999	2,46	1,20	-1,55	22,15	25,11	6,93
k = -0,3						
0,5	0,19	0,01	0,09	1,14	0,30	0,08
0,8	0,15	-0,03	-0,01	1,26	0,41	0,17
0,9	0,12	-0,11	-0,26	1,60	0,71	0,41
0,99	0,21	-0,74	-0,69	15,73	11,96	8,18
0,999	2,11	-1,82	-2,75	43,88	26,49	15,82

**Tableau 2.** Biais et REQM des quantiles estimés par les méthodes MV, LM et GML pour le modèle GEV2.**Table 2.** Bias and Root mean square error of the quantile estimators given by the Maximum likelihood, L-moment and Generalized maximum likelihood methods for the GEV2 model.

$\rho$	Bias			RMSE		
	MV	LM	GML	MV	LM	GML
k = -0,1						
0,5	-0,07	-0,07	0,02	0,33	0,33	0,25
0,8	-0,11	-0,10	0,06	0,40	0,39	0,34
0,9	-0,15	-0,14	0,08	0,53	0,52	0,46
0,99	-0,29	-0,31	0,18	3,26	2,73	1,26
0,999	-0,20	-0,43	0,31	24,63	17,07	16,36
k = -0,2						
0,5	0,17	0,17	0,08	0,68	0,67	0,17
0,8	0,13	0,13	0,06	0,86	0,84	0,46
0,9	0,09	0,11	-0,01	1,12	1,08	0,62
0,99	-0,44	-0,62	-0,19	8,36	6,31	2,87
0,999	-0,91	-1,85	-1,44	36,97	24,36	9,18
k = -0,3						
0,5	0,04	0,14	-0,20	1,32	0,87	0,88
0,8	-0,25	0,21	-0,19	1,25	1,03	1,06
0,9	-0,31	-0,38	-0,27	1,37	1,35	1,30
0,99	-1,31	-1,01	-0,76	11,61	10,77	3,10
0,999	-4,77	-3,18	-2,38	34,75	25,40	15,45

une diminution de la variance des estimateurs du maximum de vraisemblance par rapport aux résultats obtenus dans le cas du modèle GEV1. Dans la majorité des cas étudiés, la méthode GML donne les meilleurs résultats, et ce, pour les trois modèles surtout au niveau de la REQM. Pour résoudre le problème de biais dans le cas des fortes asymétries, il serait intéressant de considérer une loi *a priori* informative pour le paramètre de forme. Ceci peut être obtenu à partir de l'information régionale, tel que suggéré par MARTINS et STEDINGER (2000) dans le cas stationnaire.

## 5. PLUIE MAXIMALE ANNUELLE À LA STATION TEHACHAPI EN CALIFORNIE

Nous avons appliqué les modèles présentés ci-dessus à une série  $\underline{X} = (x_1, \dots, x_n)$  de précipitations maximales annuelles instantanées en mm, enregistrées à la station Tehachapi de la Californie (Station 048826 Latitude : 35.13 Longitude : -118.45, période 1952-2000 et le nombre d'années d'enregistrement  $n = 49$ ). La figure 2 montre la localisation géographique de la station d'étude. Située au sud de la Californie, les précipitations

de cette région sont fortement affectées par le phénomène El Niño (HASTON et MICHAELSEN, 1997).

Dans cette application du modèle GEV non stationnaire, l'objectif est d'étudier la distribution des précipitations maximales annuelles (X) en fonction de l'indice d'oscillation du sud («Southern Oscillation Index», SOI). Le coefficient de corrélation entre la variable X et SOI est de  $\rho = -0,57$  (Figure 3).

On remarque qu'il y a une corrélation négative significative entre les deux variables. À partir des figures 3 et 4, on remarque que les valeurs extrêmes de la série des précipitations correspondent aux très faibles valeurs de SOI. Les trois cas du modèle GEV non stationnaire, présentés dans la section 2 du présent travail, sont utilisés pour l'étude de la distribution de la pluie maximale annuelle avec, comme covariable, l'indice SOI.

Les paramètres des trois modèles GEV0, GEV1 et GEV2 sont estimés par la méthode GML qui, dans la majorité des cas étudiés dans la section précédente, était meilleure que la méthode du maximum de vraisemblance et la méthode des L-moments. L'estimateur GML correspond au mode de la loi *a posteriori* de chacun des paramètres. La figure 5 présente les  $N = 15\ 000$  itérations de l'algorithme de Metropolis-



Figure 2. Situation géographique de la station Tehachapi de la Californie.  
*Geographical location of the Tehachapi station in California.*

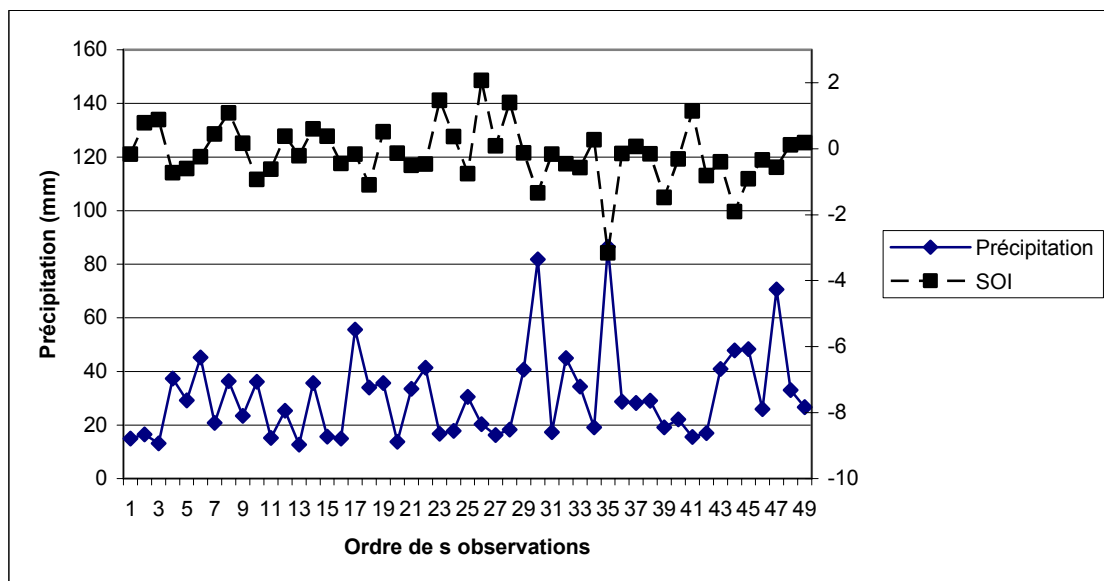


Figure 3. Séries observées des précipitations maximales annuelles et de l'indice SOI.  
*Observed values of the maximum annual precipitation and the SOI.*

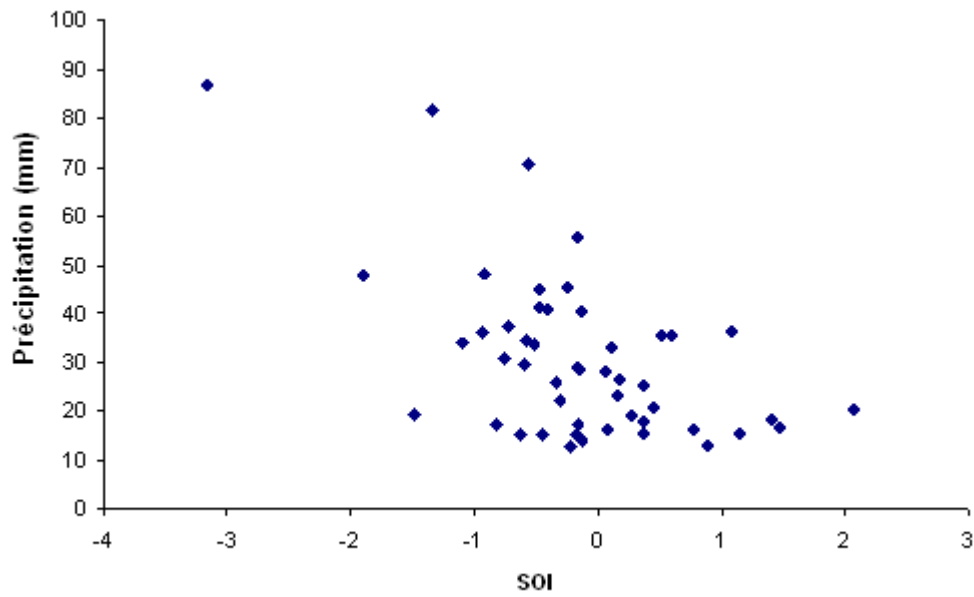


Figure 4. Nuage de pluie maximale annuelle observée en fonction de l'indice SOI correspondant.  
*Scatter plot of the maximum annual precipitation and the SOI.*

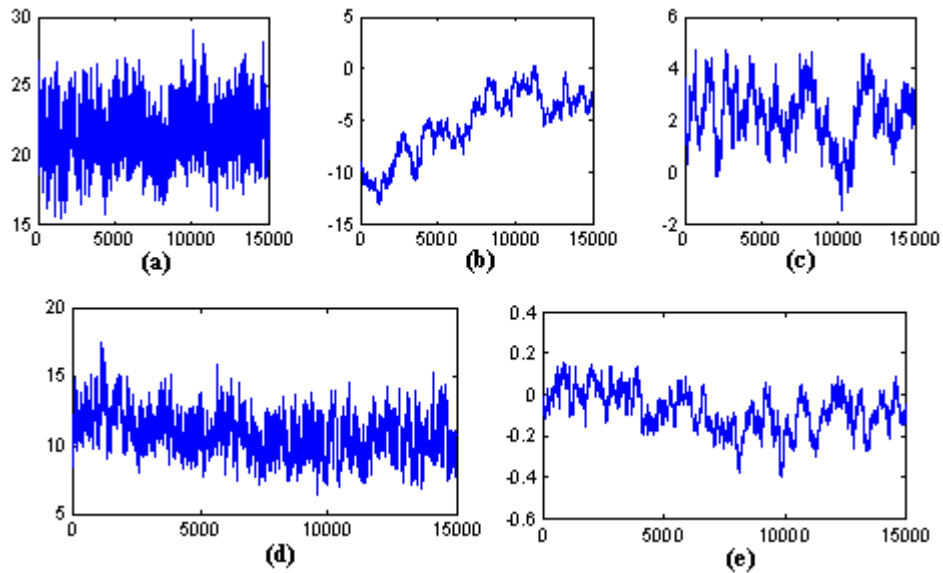


Figure 5. Itérations de l'algorithme M-H pour l'estimation des paramètres de position (a, b et c), du paramètre d'échelle (d) et de forme (e) du modèle GEV2.  
*Metropolis-Hastings Iterations to estimate the location (a, b and c), scale (d) and shape (e) parameters of the GEV2 model.*

Hastings pour l'estimation des paramètres du modèle GEV2. Après un certain nombre d'itérations qui représente le temps de chauffe, les histogrammes des derniers  $N - N_0$  itérations représentent les distributions empiriques *a posteriori* des paramètres et permettent de déterminer toutes les propriétés de ces lois *a posteriori* (Figure 6). Les estimateurs des paramètres correspondent aux modes de ces distributions empiriques et sont donnés dans le tableau 3 avec la valeur de la fonction de log-vraisemblance maximisée  $l_n^*$ .

Le modèle GEV0 est un cas particulier du modèle GEV1 qui est également un cas particulier du modèle de dépendance quadratique GEV2. Le modèle le plus général est le meilleur pour représenter la variance des données. Quand la différence entre deux modèles n'est pas évidente, il est toujours préférable de considérer le modèle le plus simple pour respecter le principe de parcimonie. Une méthode simple pour comparer la validité

d'un modèle  $M_1$  par rapport à un autre  $M_0$ , tel que  $M_0 \subset M_1$ , est d'utiliser la statistique de déviance définie par :

$$D = 2 \{l_n^*(M_1) - l_n^*(M_0)\} \quad (25)$$

où  $l_n^*$  est la fonction de log-vraisemblance maximisée de chacun des modèles. Les valeurs élevées de  $D$  indiquent que le modèle  $M_1$  explique plus la variation des données que le modèle  $M_0$ , et est donc le plus adéquat. La statistique  $D$  est distribuée suivant une loi chi-deux de paramètre  $v(\chi_v^2)$ , qui est la différence entre le nombre de paramètres du modèle  $M_1$  et celui du modèle  $M_0$ . Les valeurs de  $D$  supérieures aux quantiles de la loi chi-deux correspondent à un certain niveau de confiance, sont considérées non nulles, et le modèle  $M_1$  est meilleur dans ce cas que le modèle  $M_0$ .

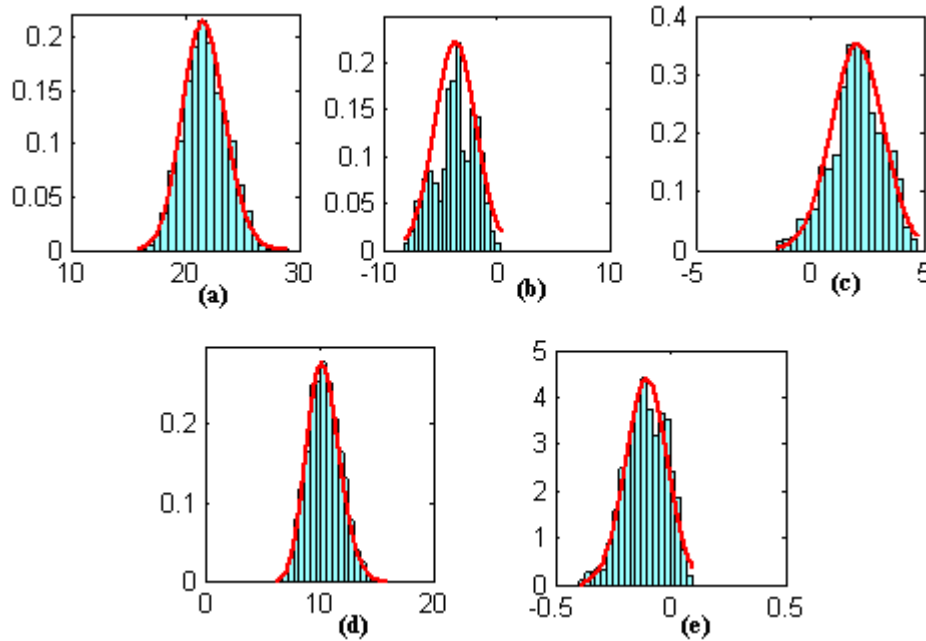


Figure 6. Distribution *a posteriori* empirique obtenue à partir des  $N - N_0$  itérations de l'algorithme MCMC des paramètres de position (a, b et c), du paramètre d'échelle (d) et de forme (e) du modèle GEV2.

*Empirical posterior distribution using the  $N - N_0$  iterations of the MCMC algorithm for the location (a, b and c), scale (d) and shape (e) parameters of the GEV2 model.*

Tableau 3. La fonction log-vraisemblance maximisée et les estimateurs des paramètres de chacun des modèles.  
Table 3. Maximized log-likelihood function and parameter estimators for all studied models.

	$l_n^*$	$\beta_1$	$\beta_2$	$\beta_3$	$\alpha$	$\kappa$
GEV0	-197,11	23,45	-	-	12,30	-0,15
GEV1	-196,16	23,01	-5,46	-	13,25	-0,12
GEV2	-192,12	21,75	-3,59	2,05	10,48	-0,10

La comparaison du modèle GEV0 au modèle GEV1 montre que la différence n'est pas significative puisque la valeur de la statistique  $D = 0,95$  est considérée significativement nulle au seuil de confiance 95 % ( $\Pr(\chi_1^2 \leq 0,95) = 0,6703$ ). Pour le modèle GEV1 par rapport au modèle GEV2 la statistique  $D = 4,04$ , ce qui montre que, dans ce cas, la différence est significative au seuil de confiance 95 % puisque  $\Pr(\chi_1^2 \leq 4,04) = 0,9556$ . Le modèle GEV2 est donc le plus adéquat pour représenter la dépendance entre le paramètre de position de la loi GEV et l'indice SOI.

La différence entre les trois modèles peut être aussi remarquée au niveau des quantiles, estimés pour chacun des modèles, conditionnellement aux valeurs de SOI. Les quantiles ainsi que leurs intervalles sont calculés à partir de l'équation 24. La figure 7 présente les médianes conditionnelles aux différentes valeurs de SOI et leurs intervalles de crédibilité au niveau 95 %, estimés par les modèles GEV0 et GEV1. Les mêmes estimateurs obtenus par le modèle GEV0 sont représentés à la figure 8 avec ceux obtenus à partir du modèle GEV2.

On remarque que la différence est plus importante pour les faibles valeurs de SOI qui correspondent aux valeurs extrêmes de la pluie maximale annuelle. En effet, la médiane estimée par le modèle quadratique GEV2 peut atteindre jusqu'à trois fois celle estimée par le modèle GEV0. La comparaison des trois modèles par le test basé sur la statistique de déviance

montre que le modèle GEV2 est plus adéquat pour représenter les données. On constate que le modèle GEV0 mène à une importante sous-estimation de la médiane dans le cas des faibles valeurs de SOI, tandis que le modèle avec tendance sous-estime la médiane pour les plus grandes valeurs de SOI.

On peut aussi reproduire la série de la pluie maximale annuelle à Tehachapi à partir de la médiane du modèle GEV2 calculée en fonction des valeurs observées de la covariable SOI. La figure 9 montre que la forme générale de la série est bien reproduite, essentiellement pour la valeur minimale de la covariable, qui correspond à la valeur maximale de la série étudiée.

## 6. CONCLUSION

Dans ce travail nous avons présenté le modèle GEV non stationnaire qui peut être utilisé dans plusieurs situations pour décrire d'une manière adéquate la variance des données. Nous nous sommes intéressés surtout au cas d'une non-stationnarité au niveau du paramètre de position de la distribution GEV. Plusieurs autres modèles statistiques permettent de représenter des variables aléatoires dans le cas d'une non-stationnarité, mais ils sont basés sur l'hypothèse de normalité. Le modèle GEV

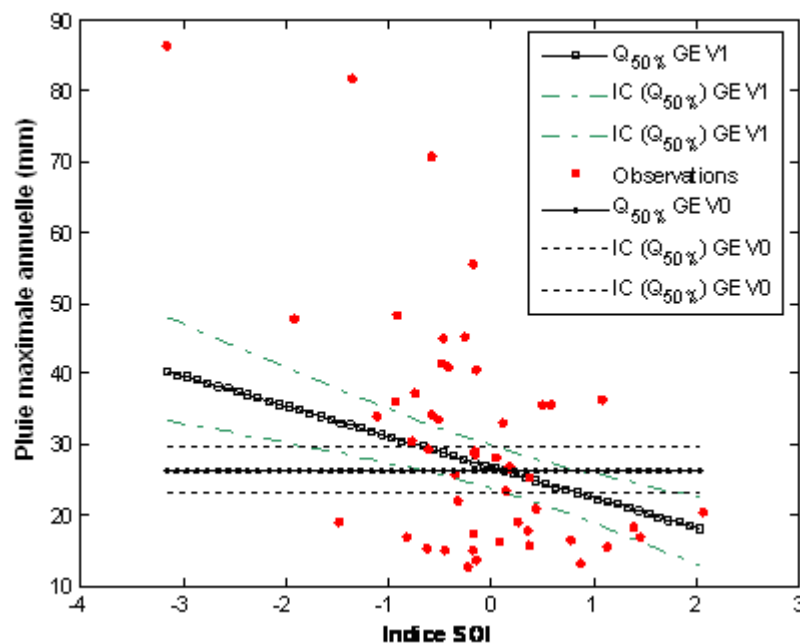


Figure 7. Estimateurs de la médiane conditionnellement aux valeurs de SOI obtenus pour les modèles GEV0 et GEV1.  
Median estimators as function of the SOI values given by the GEV0 and GEV1 models.

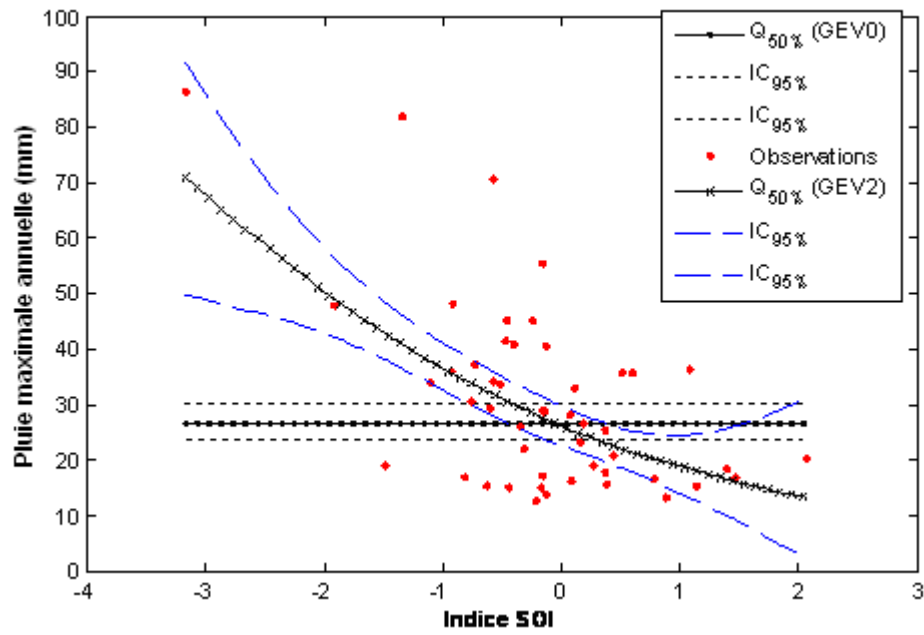


Figure 8. Estimateurs de la médiane conditionnellement aux valeurs de SOI obtenus par les modèles GEV0 et GEV2.

*Median estimators as functions of the SOI values given by the GEV0 and GEV2 models.*

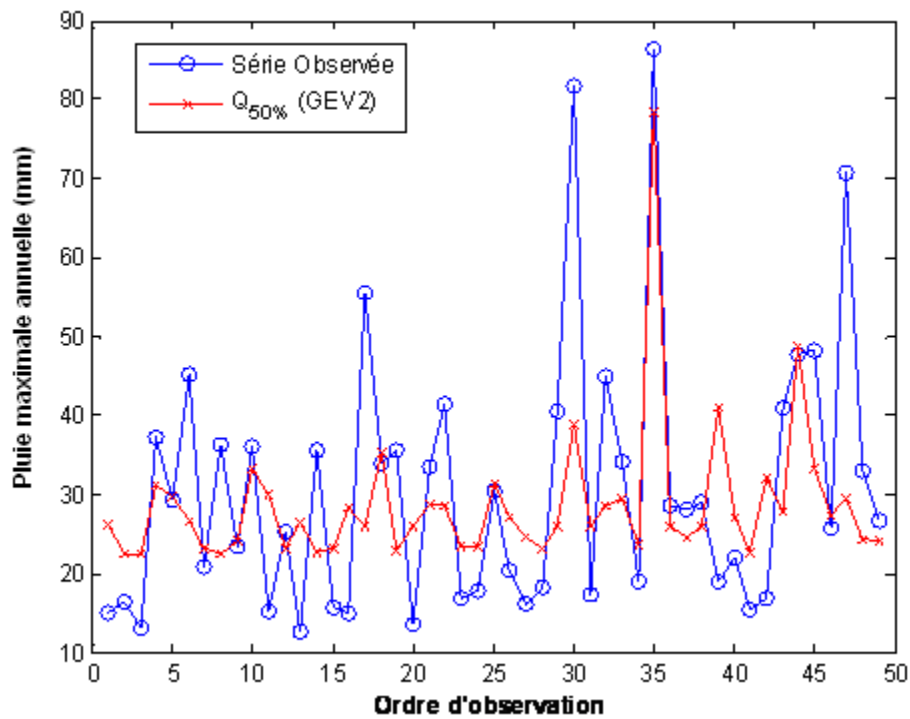


Figure 9. Estimateurs de la médiane conditionnellement aux valeurs observées de l'indice SOI obtenus par le modèle GEV2.

*Median estimators as functions of the SOI values given by the GEV2 model.*

non stationnaire est le plus adéquat pour représenter la non-stationnarité dans le cas de variables de valeurs extrêmes. Les modèles étudiés dans ce travail correspondent au cas classique, le cas d'une tendance au niveau du paramètre de position et le cas où le paramètre de position est une fonction quadratique d'une covariable.

L'estimation des paramètres de ces modèles est faite généralement par la méthode du maximum de vraisemblance. L'objectif principal de ce travail était d'introduire une généralisation de la méthode des L-moments pour l'estimation des paramètres de la GEV dans le cas d'une non-stationnarité au niveau du paramètre de position. Cette approche est comparée à la méthode du maximum de vraisemblance et la méthode du maximum de vraisemblance généralisée introduite dans le cas non stationnaire par EL ADLOUNI *et al.* (2007). Les résultats de la comparaison par simulation montrent que la méthode des L-moments est meilleure que la méthode du maximum de vraisemblance au niveau du biais et du RMSE. Par rapport à la méthode GML, l'approche basée sur les L-moments donne d'aussi bons résultats surtout pour les faibles asymétries. Le fait de considérer une loi *a priori* sur le paramètre de forme de la loi GEV mène à une réduction considérable de la REQM des quantiles.

Les trois modèles, le classique GEV0, avec tendance GEV1 et le cas de la dépendance quadratique GEV2, ont été appliqués pour l'estimation des quantiles de la pluie maximale annuelle de la station Tehachapi de la Californie, conditionnellement à l'indice climatique SOI. La comparaison de ces modèles, avec un test basé sur la statistique de déviance, montre que le modèle GEV2 est le plus adéquat pour représenter la variance des données. Cet exemple montre l'importance de considérer le modèle non stationnaire pour tenir compte des dépendances entre la variable étudiée et d'autres covariables pour améliorer la précision des estimateurs. Nous avons remarqué que l'estimateur de la médiane conditionnellement à une faible valeur de l'indice SOI pour le modèle GEV2, peut atteindre trois fois la valeur de l'estimation de la médiane par le modèle GEV0.

Ces modèles constituent un outil efficace pour tenir compte des dépendances entre les variables aléatoires de valeurs extrêmes. Il s'agit d'une généralisation des modèles classiques, basés sur l'hypothèse de normalité, qui, dans le cas des études hydrologiques, permet d'introduire l'effet des changements climatiques sur l'évolution des valeurs des paramètres. En effet, la non-stationnarité peut être reliée non seulement au paramètre de position mais aussi aux deux autres paramètres : d'échelle et de forme. Le cas le plus intéressant est celui de la dépendance des paramètres de position et d'échelle du temps ou de covariables car une tendance est souvent accompagnée d'un changement au niveau de la variance.

## RÉFÉRENCES BIBLIOGRAPHIQUES

- CHEN, H. et R. RAO (2002). Testing hydrologic time series for stationarity. *J. Hydrol. Eng.*, 7, 129-136.
- CLARKE, R.T. (2002). Estimating trends in data from the Weibull and a generalized extreme value distribution. *Water Resour. Res.*, 38, 25-1- 25-10.
- COLES, G.S. (2001). *An introduction to statistical modeling of extreme values*, SPRINGER (éditeur), 208 p.
- EL ADLOUNI, S., A.-C. FAVRE et B. BOBÉE (2006). Comparison of methodologies to assess the convergence of Markov Chain Monte-Carlo Methods. *Comput. Stat. Data Anal.*, 50, 2685-2701.
- EL ADLOUNI, S., T.B.M.J. OUARDA, X. ZHANG, R. ROY. et B. BOBÉE (2007). Generalized maximum likelihood estimators of the non-stationary GEV model parameters. *Water Resour. Res.*, 43, W03410, doi:10.1029/2005WR004545.
- FISHER, R.A. et L.H.C. TIPPETT (1928). *Limiting forms of the frequency distribution of the largest or smallest member of a sample*. Dans : Proceedings of the Cambridge Philosophical Society, 24, pp. 180-190.
- GREENWOOD, J.A., J.M. LANDWEHR, N.C. MATALAS et J.R. WALLIS (1979). Probability weighted moments: definition and relation to parameters of several distributions expressible in inverse form. *Water Resour. Res.*, 15, 1049-1054.
- HAMILTON, J. (1991). A quasi-Bayesian approach to estimating parameters for mixtures of normal distributions. *J. Bus. Econom. Statist.*, 9, 27-39.
- HASTINGS, W. (1970). Monte Carlo sampling methods using Markov Chains and their applications. *Biometrika*, 57, 97-109.
- HASTON, L. et J. MICHAELSEN (1997). Spatial and temporal variability of southern California precipitation over the last 400 yr and relationships to atmospheric circulation patterns. *J. Clim.*, 10, 1836-1852.
- HOSKING, J.R.M. (1990). L-Moments: Analysis and estimation of distributions using linear combinations of order statistics. *J. Royal Stat. Soc.*, 52, 105-124.



- INTERGOVERNMENTAL PANEL FOR CLIMATE CHANGE [IPCC] (2001). *Climate Change 2001: Impacts, adaptation and vulnerability*, Cambridge Univ. Press, New York.
- JENKINSON, A.F. (1955). The frequency distribution of the annual maximum (or minimum) of meteorological elements. *Quart. J. Roy. Meteor. Soc.*, 81, 158-171.
- KATZ, R.W. (1999). Extreme value theory for precipitation: sensitivity analysis for climate change. *Adv. Water Resour.*, 23, 133-139.
- KATZ, R.W., M.B. PARLANGE et P. NAVEAU (2002). Statistics of extremes in hydrology. *Adv. Water Resour.*, 25, 1287-1304.
- KHARIN, V.V. et F.W. ZWIERS (2005). Estimating extremes in transient climate change simulations. *J. Clim.*, 18, 1156-1173.
- METROPOLIS, N., A. ROSENBLUTH, M. ROSENBLUTH, A. TELLER et E. TELLER (1953). Equations of state calculations by fast computing machines. *J. Chem. Phys.*, 21, 1087-1092.
- MARTINS, E.S. et J.R. STEDINGER (2000). Generalized maximum likelihood GEV quantile estimators for hydrologic data. *Water Resour. Res.*, 36, 737-744.
- MORRISON, J.E. et J.A. SMITH (2003). Stochastic modeling of flood peaks using the Generalized Extreme Value (GEV) distribution, *Water Resour. Res.*, 38, 41-1 – 41-12.
- ÖNÖZ, B. et M. BAYAZIT (2003). The power of statistical tests for trend detection. *Turkish J. Eng. Env. Sci.*, 27, 247-251.
- OUARDA, T.B.M.J., P.F. RASMUSSEN, J.-F. CANTIN, B. BOBÉE, R. LAURENCE, V.D. HOANG et G. BARABÉ (1999). Identification d'un réseau hydrométrique pour le suivi des changements climatiques : Application à la province de Québec, *Rev. Sci. Eau*, 12, 425-448.
- SCARF, P.A. (1992). Estimation for a four parameter generalized extreme value distribution, *Comm. Statist. A Theory Methods*, 21, 2185-2201.
- SMITH, R.L. (1985). Maximum likelihood estimation in a class of non-regular cases. *Biometrika* 72, 67-92.
- SANKARASUBRAMANIAN, A. et U. LALL (2003). Flood quantiles in a changing climate: Seasonal forecasts and causal relations. *Water Resour. Res.*, 39, art. no. 1134.
- STEDINGER, J.R., R.M. VOGEL et E. FOUFOULA-GEORGIO (1993). *Frequency analysis of extreme events*. Dans : Handbook of Hydrology, D.R. MAIDMENT (éditeur), McGraw Hill Inc., Chapitre 18, pp. 1-66.
- VENKATARAMAN, S. (1997). *Value at risk for a mixture of normal distributions: The use of quasi-Bayesian estimation technique*. Economic Perspectives. Federal Reserve Bank of Chicago, pp. 2-13.
- WANG, X.L., F.W. ZWIERS et V. SWAIL (2004). North Atlantic Ocean wave climate scenarios for the 21<sup>st</sup> century. *J. Clim.*, 17, 2368-2383.
- YUE, S., P. PILON et G. CAVADIAS (2002). Power of the Mann-Kendall and Spearman's rho tests for detecting monotonic trends in hydrologic series. *J. Hydrol.*, 259, 254-271.
- ZHANG, X., K.D. HARVEY, W.D. HOGG et T.R. YUZYK (2001). Trends in Canadian streamflow. *Water Resour. Res.*, 37, 987-999.