

Université du Québec  
INRS-Eau, Terre et Environnement

# DÉVELOPPEMENT DU MODÈLE LOG-NORMAL NON STATIONNAIRE ET APPLICATION À L'ESTIMATION DES EXTRÊMES

Par  
Ilham Aissaoui Fqayeh

Mémoire présenté pour l'obtention  
du grade de Maîtrise ès sciences (M.Sc.)  
en Sciences de l'eau et de l'environnement

Jury d'évaluation

Examineur externe

Pr. Saad Bennis  
École de Technologie Supérieure

Examineur interne

Pr. Alain Maillot  
INRS-Eau, Terre et Environnement

Codirecteurs de recherche

Dr. Salah El Adlouni  
Pr. André St-Hilaire  
INRS-Eau, Terre et Environnement

Directeur de recherche

Pr. Taha B.M.J Ouarda  
INRS-Eau, Terre et Environnement



# TABLE DES MATIÈRES

<b>RÉSUMÉ</b> .....	<b>V</b>
<b>SECTION 1 : SYNTHÈSE</b> .....	<b>VII</b>
<b>REMERCIEMENTS</b> .....	<b>VII</b>
<b>INTRODUCTION</b> .....	<b>1</b>
<b>MISE EN CONTEXTE</b> .....	<b>1</b>
<b>PROBLÉMATIQUE</b> .....	<b>1</b>
<b>OBJECTIFS DU MÉMOIRE</b> .....	<b>3</b>
<b>CHAPITRE 1 : CONTRIBUTION DE L'ÉTUDIANTE</b> .....	<b>5</b>
<b>CHAPITRE 2 : SITUATION DE LA CONTRIBUTION</b> .....	<b>7</b>
<b>RÉFÉRENCES</b> .....	<b>7</b>
<b>SECTION 2 : ARTICLE 1</b> .....	<b>11</b>
<b>1 INTRODUCTION</b> .....	<b>13</b>
<b>2 MODÈLE LOG-NORMAL NON-STATIONNAIRE</b> .....	<b>15</b>
2.1 Caractéristiques statistiques .....	15
2.1.1 Distribution Log-normale .....	15
2.1.2 Modèle Log-normal Non-stationnaire .....	17
2.2 Estimation des paramètres par la méthode du maximum de vraisemblance.....	19
2.3 Estimation des quantiles.....	21
<b>3 COMPARAISON DES MODÈLES LN ET GEV NON-STATIONNAIRES</b> .....	<b>22</b>
3.1 Plan de simulation .....	22
3.2 Résultats .....	25
<b>4 APPLICATION : PRÉCIPITATIONS À LA STATION TEHACHAPI</b> .....	<b>29</b>
<b>5. CONCLUSIONS</b> .....	<b>34</b>
<b>RÉFÉRENCES</b> .....	<b>35</b>
<b>ANNEXE A</b> .....	<b>49</b>
<b>ANNEXE B</b> .....	<b>51</b>
<b>SECTION 3 : ARTICLE 2</b> .....	<b>53</b>
<b>1 INTRODUCTION</b> .....	<b>53</b>

<b>2</b>	<b>RAPPEL SUR LE MODÈLE LOG-NORMAL NON-STATIONNAIRE.....</b>	<b>53</b>
2.1.	Test sur la tendance:.....	55
2.2.	Le modèle Log-normal non-stationnaire.....	56
<b>3</b>	<b>CARACTÉRISTIQUES STATISTIQUES DES SÉRIES ÉTUDIÉES.....</b>	<b>57</b>
3.1	Résultats du test de Mann-Kendall .....	57
3.2	Caractéristiques statistiques .....	58
<b>4</b>	<b>ESTIMATION DES PARAMÈTRES ET DE LA MÉDIANE .....</b>	<b>58</b>
<b>5</b>	<b>DISCUSSION DES RÉSULTATS.....</b>	<b>59</b>
	<b>RÉFÉRENCES.....</b>	<b>61</b>
	<b>LISTE DE TABLEAUX .....</b>	<b>65</b>
	<b>LISTE DE FIGURES.....</b>	<b>65</b>
	<b>ANNEXE A.....</b>	<b>73</b>
	<b>ANNEXE B.....</b>	<b>77</b>
	<b>ANNEXE C.....</b>	<b>81</b>
	<b>SECTION 4 : CONCLUSION .....</b>	<b>83</b>

## RÉSUMÉ

La connaissance des débits de crue est nécessaire pour la conception des aménagements des cours d'eau, le dimensionnement des ouvrages hydrauliques et la protection des zones urbaines. Des études statistiques spécifiques sont nécessaires pour la prédétermination des débits de crue. L'analyse fréquentielle est un des outils privilégiés pour l'estimation des débits de crue. Les approches classiques d'analyse fréquentielle nécessitent l'hypothèse de stationnarité des séries d'observations. Cette hypothèse est de plus en plus mise en doute dans le contexte des changements climatiques observés.

L'objectif principal de cette étude est de présenter le modèle Log-normal non-stationnaire pour l'estimation des séries non-stationnaires et de le comparer au modèle GEV non-stationnaire. L'estimation des paramètres est effectuée par la méthode du maximum de vraisemblance. Ce genre de modèle permet d'introduire des covariables au niveau du premier paramètre de la loi Log-normale ou le temps dans le cas de tendance temporelle. Trois formes différentes du modèle Log-normal non stationnaire sont investiguées: modèle classique, modèle avec tendance linéaire, et modèle avec tendance quadratique. Ces modèles ont été comparés, par simulation de Monte Carlo, aux modèles généralisés des valeurs extrêmes (GEV) non-stationnaires, qui ont été bien étudiés dans la littérature. Les modèles non-stationnaires, d'une manière générale, représentent des outils efficaces pour tenir compte de l'effet des changements climatiques en modélisant l'évolution des valeurs des paramètres en fonction du temps ou d'autres covariables. L'utilité de ces modèles est illustrée par l'étude de l'effet d'un indice climatique (Southern Oscillation Index, SOI) sur des données hydro-climatiques de la Californie, USA. L'effet de l'indice SOI sur les précipitations maximales annuelles, enregistrées à la station Tehachapi de la Californie est étudié

dans le premier article. Le deuxième article traite de l'effet de la tendance sur des données de débit de pointe dans plusieurs stations à travers le monde.

## **SECTION 1 : SYNTHÈSE**

### **REMERCIEMENTS**

J'exprime mes profonds remerciements à mon directeur de thèse, le professeur Taha B.M.J Ouarda pour l'aide compétente qu'il m'a apportée, pour sa patience et son encouragement à finir ce travail commencé il y a longtemps. Son oeil critique m'a été très précieux pour structurer le travail et pour améliorer la qualité des différentes sections.

Mes deux co-directeurs de thèse Salah El Adlouni et André St-Hilaire qui ont beaucoup contribué à mettre en forme mon français et mon anglais très approximatifs. Sans eux, j'aurais sans doute été découragée par ces langues que j'aime bien.

Je remercie mon mari, Abdelkader, pour l'amour qu'il me porte et pour son soutien et sa patience pendant que je travaillais souvent presque toute la journée et parfois le soir aussi.

D'autres personnes m'ont encouragé à finir ce travail qui sont : ma mère que j'aime beaucoup, mon père, mes sœurs et mes frères.

Mon adorable bébé Adam qui a 11 mois maintenant, et qui était très gentil avec moi durant la période de grossesse, il m'a laissé travailler sans me poser des problèmes connus chez la plupart des femmes enceintes.

"Last but not least", j'exprime ma gratitude aux membres de mon jury de thèse: Taha B.M.J Ouarda, Salah El Adlouni, André St-Hilaire, Saad Bennis et Alain Maillot.



# INTRODUCTION

## Mise en contexte

Ce mémoire « par Article » se divise en deux parties. La première partie qui représente le premier article, fait état de l'introduction du modèle Log-normal non-stationnaire, et de la méthode du maximum de vraisemblance pour l'estimation des paramètres. Ce modèle a été ensuite comparé au modèle GEV non-stationnaire pour évaluer les erreurs liées à la méthode d'estimation et celles correspondant à un mauvais choix du modèle. On y décrit la problématique, les objectifs, la méthodologie et les résultats de l'application de ces deux familles de modèles à des précipitations maximales annuelles journalières, enregistrées à la station Tehachapi de la Californie.

La deuxième partie du mémoire est constituée d'un deuxième article où on applique la nouvelle approche aux débits de pointe dans différentes stations répartis à travers le monde.

## Problématique

Les valeurs extrêmes sont souvent les maximums 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 (GEV). Les conditions théoriques de l'utilisation de la loi GEV ne sont pas souvent vérifiées et d'autres lois sont considérées pour représenter les différentes formes d'extrêmes observés. Notons également que les approches classiques d'analyse fréquentielle qui supposent la stationnarité des séries d'observations, ainsi que l'indépendance et l'homogénéité ne sont pas toujours vérifiées. En effet, en hydrologie la loi de probabilité des crues peut changer avec le temps (existence de non-stationnarité) comme conséquence des activités humaines ou à cause des changements climatiques (El Adlouni *et al.* 2007, Zhang *et al.* 2001). Il faudrait donc développer des outils qui tiennent compte de la non-stationnarité des séries

hydro-climatiques ou de dépendance d'autres covariables. Ce genre de modèles permet d'inclure l'effet de différentes covariables sur la variabilité et l'évolution de la série observée.

Dans le cadre d'une analyse fréquentielle, les hypothèses d'indépendance, d'homogénéité et de stationnarité des séries observées doivent être vérifiées. Plusieurs tests permettent de vérifier ces conditions. On s'intéresse principalement, dans le cadre de cette étude, à l'hypothèse de non-stationnarité. Plusieurs tests permettent de tester cette hypothèse comme par exemple le test de Mann, Kruskal-Wallis (Faucher *et al.*, 1997) et Mann-Kendall (Önöz et Bayazit, 2003). Récemment de nouveaux tests ont été développés tel que le test de segmentation (Chen et Rao, 2002), et le test de Kwiatkowski-Phillips-Schmidt-Shin (le test KPSS) (Hobijn, 2004).

L'estimation des quantiles en analyse fréquentielle dépend des propriétés des estimateurs des paramètres de la distribution à ajuster. Les méthodes d'estimation les plus utilisées sont la méthode des moments de probabilités pondérées (PWM) et la méthode des L-moments (LM) (Hosking 1990) en raison de leur simplicité pour la majorité des distributions. Cependant, dans un cadre non-stationnaire, les estimateurs des L-moments ne peuvent pas être obtenus d'une manière directe. On considère souvent l'estimation par la méthode du maximum de vraisemblance (ML) qui permet, par le biais de la résolution numérique du système d'équations, d'estimer les paramètres même dans des cas très complexes (Coles, 2001).

Les travaux sur les modèles d'analyse fréquentielle dans un cadre non-stationnaires (ou en présence de dépendance de covariables) sont liés au modèle GEV (Coles 2001 et Clarke 2002a,b). En effet, la loi GEV est la loi la plus utilisée en hydrologie pour la modélisation des extrêmes. C'est dû principalement aux propriétés théoriques et à la simplicité de la fonction de densité de la loi GEV et sa fonction quantile. La loi GEV a été, également jugée dans plusieurs études comme adéquate pour l'ajustement des séries des maximums (Faucher *et al.* 1997). Dans

d'autres régions, d'autres distributions ont été suggérées pour l'étude des séries hydrométéorologiques : Log-normale (LN) en Chine, Log-Pearson type III (LPIII) aux États-Unis d'Amérique (USWRC, 1982), ou la loi Log-normale à deux paramètres (LN) au Québec (Kouider, 2003). Notons que la loi LN a un comportement au niveau des extrêmes qui est très semblable à celui des distributions à queues lourdes (Ouarda *et al.*, 1994). Il est souvent très difficile de discriminer entre la loi LN et les distributions de type puissance (comme par exemple la loi GEV et la loi Log-Pearson type III).

### **Objectifs du mémoire**

Dans le premier article on présente le modèle Log-normal non-stationnaire à deux paramètres. Ce modèle s'ajuste bien à plusieurs variables hydrométéorologiques. Il est également important de disposer de modèles non-stationnaires de distributions à deux paramètres à cause du principe de parcimonie. En effet, le nombre des paramètres augmente avec le nombre de covariables d'où l'avantage de considérer une distribution avec le moins de paramètres possible, surtout en hydrologie où la taille des séries d'observations est souvent faible.

Le deuxième objectif de l'article 1 est de comparer, sur la base de simulations Monte Carlo, les propriétés des estimateurs des quantiles obtenus par la méthode du maximum de vraisemblance à partir des modèles LN non-stationnaires et GEV non-stationnaires pour évaluer les erreurs liées à la taille de l'échantillon et à l'efficacité de la méthode d'estimation ainsi que les erreurs générées par un mauvais choix du modèle. Une application des deux familles de modèles non-stationnaires LN et GEV à des données de précipitations maximales annuelles conditionnellement à l'indice climatique SOI a permis de comparer leurs résultats, et d'illustrer la flexibilité de ces modèles pour l'ajustement des séries hydrométéorologiques.

L'objectif principal de l'article 2 est d'appliquer le modèle Log-normal non stationnaire à des séries de débit maximum annuel de 17 stations tirées du site de l'UNESCO et qui correspondent à différentes rivières du monde. Les critères de sélection de ces séries sont : (1) la taille de la série (supérieure à 70 ans d'observations), (2) les tests d'ajustement montrent que la loi Log-normale représente bien la série étudiée. On considère ensuite le modèle Log-normal non-stationnaire pour l'estimation des tendances temporelles qui pourraient exister dans ces séries de débit de crue.

## **CHAPITRE 1 : CONTRIBUTION DE L'ÉTUDIANTE**

La problématique et l'approche globale ont été proposées par mon directeur de recherche. J'ai réalisé une revue de littérature sur les modèles non-stationnaires et les tests de détection de non-stationnarité durant l'hiver 2005. Les résultats de cette revue sont présentés dans le chapitre «Introduction» de l'article 1. J'ai ensuite travaillé, sous la supervision de mes directeurs de recherche, au développement des équations mathématiques pour l'estimation des paramètres par la méthode du maximum de vraisemblance (ML) pour le modèle Log-normal non-stationnaire et la comparaison par simulation des modèles LN et GEV non-stationnaires. Ces développements sont présentés respectivement dans les sections 2 et 3 de l'article 1. J'ai ensuite analysé les résultats des simulations avec la contribution des mes directeurs de recherche.

J'ai colligé les banques de données des débits maximums annuels de plusieurs rivières du monde à partir du site de l'UNESCO pour l'étude de la non-stationnarité et l'application des modèles étudiés. Les résultats de cette application sont présentés dans l'article 2.

J'ai entièrement programmé les codes Matlab pour tester la stationnarité des séries étudiées, les codes pour les simulations Monte Carlo ainsi que les codes pour le calcul des estimateurs des paramètres et des quantiles sous la supervision du Dr. El Adlouni.

J'ai rédigé les deux articles avec la contribution de mes directeurs de recherche. J'ai ensuite rédigé le présent mémoire qui a été corrigé par mes directeurs de recherche.



## CHAPITRE 2 : SITUATION DE LA CONTRIBUTION

Dans le premier travail (article1), après avoir présenté les modèles LN non-stationnaires les résultats des simulations ont permis d'évaluer les erreurs liées à la taille de l'échantillon et la méthode d'estimation ainsi que les erreurs qui sont dues à un mauvais choix du modèle.

La contribution originale du présent travail est de présenter un nouveau modèle pour l'ajustement des séries à des lois de probabilité dont les paramètres sont fonctions de covariables. Le choix du modèle Log-normal est justifié par deux principales raisons : (1) la loi Log-normale est très utilisée pour la modélisation des séries hydrométéorologiques et possède des propriétés très intéressantes pour l'ajustement des extrêmes (queue de distribution); (2) elle a seulement deux paramètres, ce qui lui permet d'avoir une flexibilité pour la modélisation d'une large gamme de variation et d'asymétrie en respectant le principe de parcimonie. Une application qui illustre la flexibilité et l'efficacité du modèle proposé et de la nouvelle approche pour l'estimation des paramètres est aussi présentée.

Dans l'article2 on étudie le débit de pointe de 17 stations tirées du site de l'UNESCO et qui correspondent à différentes rivières du monde. Les critères de sélection de ces séries sont : (1) la taille de la série (supérieure à 70 ans d'observations), (2) les tests d'ajustement montrent que la loi Log-normale semble être adéquate pour représenter la série étudiée. Le modèle Log-normal non-stationnaire est considéré pour l'estimation des débits de crue, pour les 17 stations. Ensuite, le choix du modèle le plus adéquat (parmi les trois différents modèles utilisés : classique et deux types de modèle non-stationnaire) pour chaque station est effectué en utilisant la statistique de déviance.



## RÉFÉRENCES

- Clarke, R. T. (2002a). Estimating time trends in Gumbel-distributed data by means of generalized linear models, *Water Resources Research*, 38(7), 1111, doi:10.1029/2001WR000917.
- Clarke, R. T. (2002b), Estimating trends in data from the Weibull and a generalized extreme value distribution, *Water Resources Research*, 38(6), 1089, doi:10.1029/2001WR000575.
- Coles, G.S. (2001). *An Introduction to Statistical Modeling of Extreme Value*, Springer, 208 p.
- El Adlouni S., Ouarda T. B.M.J., Zhang X., Roy R. et Bobée B. (2007). Generalized maximum likelihood estimators for the non-stationary GEV model. *Water Resources Research*, sous presse.
- Faucher, D., Ouarda, T.B.M.J. et B. Bobée (1997). *Revue bibliographique des tests de stationnarité*. INRS-Eau, rapport de Recherche no R-499, 66 pages.
- Hosking, J.R.M. (1990). L-Moments: Analysis and Estimation of Distributions Using Linear Combinations of Order Statistics. *Newspaper of the Royal Statistical Society*, 52:105-124.
- Katz R.W., Parlange M.B. et Naveau P. (2002). Statistics of extremes in hydrology, *Advances in Water Resources*, vol. 25, pp. 1287-1304.
- Kouider A. (2003). *Analyse fréquentielle locale des crues au Québec*. M. Sc. Thesis, INRS-ETE, University of Quebec, Canada.
- Martins E.S. et Stedinger J.R. (2000). Generalized Maximum Likelihood GEV Quantile Estimators for Hydrologic Data. *Water Resources Research*, vol. 36. No. (3). pp.737-744.
- USWRC (1982). *Guidelines for Determining Flood Flow Frequential*. Bulletin 17B, Hydrology Committee, (U.S. Water Resources Council): Washington, DC.
- Zhang X., K. D. Harvey, W. D. Hogg, et T. R. Yuzyk (2001). Trends in Canadian streamflow. *Water Resources Research*, Vol. 37, No. 4, pp. 987–999.



## **SECTION 2 : ARTICLE 1**

### **Développement du modèle Log-normal non-stationnaire et comparaison avec le modèle GEV non-stationnaire**

I. Aissaoui-Fqayeh

S. El Adlouni

T. B.M.J. Ouarda

A. St-Hilaire

Chaire en hydrologie statistique (Hydro-Québec / CRSNG)

Chaire du Canada en estimation des variables hydrologiques

INRS-ETE, Université du Québec

490, rue de la Couronne (Québec), Canada G1K 9A9

Soumis au Journal des Sciences Hydrologiques

Mars 2007



## 1 INTRODUCTION

L'étude des processus et événements hydrologiques extrêmes à l'échelle locale et régionale passe par la mesure et l'analyse de différentes variables hydro-climatiques (précipitations, températures, débits, etc.). Cela nécessite des outils statistiques adaptés qui tiennent compte de l'interaction entre les différentes variables. Les approches classiques d'analyse fréquentielle supposent la stationnarité des séries d'observations, ainsi que l'indépendance et l'homogénéité. Autrement dit, les observations doivent être indépendantes et identiquement distribuées (i.i.d). En hydrologie, la loi de probabilité des crues peut changer avec le temps (existence de non-stationnarité) comme conséquence des activités humaines ou à cause des changements climatiques (El Adlouni *et al.*, 2007, Zhang *et al.*, 2001). Il est donc nécessaire de développer des approches d'analyse fréquentielle qui tiennent compte de la non-stationnarité des séries des données hydro-climatiques. Ce genre de modèles permet d'inclure l'effet de différentes covariables sur la variabilité et l'évolution de la série observée.

Avant de procéder à une analyse fréquentielle il faut vérifier les hypothèses d'homogénéité et de stationnarité des séries observées. Plusieurs tests permettent de vérifier ces hypothèses. Dans ce travail on s'intéresse au problème de non-stationnarité. Pour tester cette dernière on utilise les tests de Mann, Kruskal-Wallis (Faucher *et al.*, 1997) et Mann-Kendall (Önöz et Bayazit, 2003). Récemment de nouveaux tests ont été développés tel que le test de segmentation (Chen et Rao, 2002), et le test de Kwiatkowski-Phillips-Schmidt-Shin (le test KPSS) (Hobijn, 2004).

En analyse fréquentielle, plusieurs méthodes ont été développées pour l'estimation des paramètres des distributions d'intérêt. La méthode des moments de probabilités pondérées (PWM) et la méthode des L-moments (LM) (Hosking 1990) sont parmi les méthodes les plus

utilisées, en raison de leur simplicité pour la majorité des distributions, de plus la méthode des L-moments (LM) introduit un faible biais pour les faibles échantillons. Dans le cas non-stationnaire (existence d'une non-stationnarité temporelle ou de dépendance de covariables) les estimateurs des L-moments sont compliqués à implémenter d'une manière explicite. Dans ce cas, on considère souvent l'estimation par la méthode du maximum de vraisemblance (ML) qui permet, par le biais de la résolution numérique du système d'équations, d'estimer les paramètres même dans des cas très complexes (Coles, 2001). El Adlouni *et al.*, (2007) ont donné une extension de la méthode du maximum de vraisemblance généralisée (GML) introduite par Martins et Stedinger (2000), au cas de présence de covariables pour le modèle GEV non-stationnaire. La méthode GML permet d'intégrer une loi a priori sur le paramètre de forme de la loi GEV et de réduire l'espace des solutions de la méthode du ML à des valeurs admissibles dans le cas des séries hydrométéorologiques.

La majorité des travaux sur les modèles non-stationnaires sont liés au modèle généralisé des valeurs extrêmes (GEV) (Coles 2001 et Clarke 2002a,b). Katz *et al.* (2002), ont utilisé ce modèle pour traiter le cadre local ainsi que régional. En effet, la loi GEV est très utilisée en hydrologie et a été jugée dans plusieurs études comme adéquate pour l'ajustement des séries des maximums (Faucher *et al.*, 1997). Cependant dans certaines régions, d'autres distributions ont été suggérées pour l'étude des séries hydrométéorologiques : Log-normale (LN) en Chine, Log-Pearson type III (LPIII) aux États-Unis d'Amérique (USWRC, 1982), ou la loi Log-normale à deux paramètres (LN) au Québec (Kouider *et al.*, 2002). Notons que la loi LN a un comportement au niveau des extrêmes très semblable à celui des distributions à queues lourdes (heavy tailed distributions) (Ouarda *et al.* 1994). Il est souvent très difficile de discriminer entre la loi LN et les distributions de type puissance (comme par exemple la loi GEV et la loi Log-Pearson type III).

Dans cette étude on présente le modèle Log-normal non-stationnaire à deux paramètres. Ce modèle, comme il a été déjà mentionné, s'ajuste bien à plusieurs séries de données hydrologiques et respecte le principe de parcimonie. En effet, le nombre des paramètres augmente avec le nombre de covariables d'où l'avantage de considérer une distribution avec le moins de paramètres possible, surtout en hydrologie où la taille des séries d'observations est souvent faible. Le deuxième objectif est de comparer, sur la base de simulations Monte Carlo, les propriétés des estimateurs de quantiles par la méthode du maximum de vraisemblance obtenus pour les modèles LN non-stationnaire et GEV non-stationnaire.

Après avoir rappelé certaines propriétés statistiques de la loi Log-normale on présente le modèle Log-normal non-stationnaire et certains des cas particuliers (sous trois formes particulières) avec la méthode du maximum de vraisemblance pour l'estimation des paramètres. Nous présentons ensuite dans la section 3 une comparaison par simulation Monte Carlo entre le modèle LN non-stationnaire et le modèle GEV non-stationnaire. Une application qui illustre la flexibilité et l'efficacité du modèle proposé sera présentée dans la section 4. Les conclusions et la discussion générale sont présentées dans la section 5.

## **2 MODÈLE LOG-NORMAL NON-STATIONNAIRE**

### **2.1 Caractéristiques statistiques**

#### **2.1.1 Distribution Log-normale**

On présente ci-dessous, quelques propriétés statistiques de la loi Log-normale (Leurent, 1998). Une variable  $X$  est distribuée suivant une loi Log-normale si son logarithme suit une loi normale.

Soient  $\mu$  et  $\sigma$ , la moyenne et l'écart-type de la transformée logarithmique,  $\log(X)$ . La variable  $Z = (\log(X) - \mu) / \sigma$  est distribuée suivant une loi normale centrée réduite. Soit  $\Phi$  la fonction de répartition de la loi normale centrée réduite et  $F$  celle de la loi Log-normale. On a donc les égalités suivantes :

$$F(x_0) = pr(\{X \leq x_0\}) = pr(\{Z \leq z_0\}) = \Phi(z_0) = \Phi((\ln(x_0) - \mu) / \sigma) \quad (1)$$

$$F^{-1}(p) = \exp(\mu + \sigma \cdot \Phi^{-1}(p)) \quad (2)$$

La fonction de répartition peut être écrite sous la forme :

$$F(x_0) = \int_0^{x_0} \frac{1}{\sigma x \sqrt{2\pi}} \exp\left(-\frac{1}{2} \left(\frac{\log(x) - \mu}{\sigma}\right)^2\right) dx \quad (3)$$

La fonction de densité de probabilité  $f(x)$  est donnée par :

$$f(x) = \frac{1}{\sigma x \sqrt{2\pi}} \exp\left(-\frac{1}{2} \left(\frac{\log(x) - \mu}{\sigma}\right)^2\right), \quad (4)$$

où  $x > 0$ ,  $\sigma > 0$  et  $\mu$  est un réel.

Les moments d'ordre 1 et d'ordre 2 sont donnés par:

$$E[X] = \exp(\mu + \sigma^2 / 2) \quad (5)$$

$$\begin{aligned} E[X^2] &= \exp(2\mu + 2\sigma^2) \\ &= (E[X])^2 \times \exp(\sigma^2) \end{aligned} \quad (6)$$

La variance s'écrit donc de la façon suivante:

$$\sigma_x^2 = \exp(2\mu + \sigma^2) \cdot (\exp(\sigma^2) - 1) = E[X]^2 \cdot (\exp(\sigma^2) - 1) \quad (7)$$

Le coefficient de variation ( $C_v$ ) et le coefficient d'asymétrie ( $C_s$ ) sont donnés par :

$$C_v = (e^{\sigma^2} - 1)^{\frac{1}{2}} \quad \text{et} \quad C_s = (e^{\sigma^2} - 1)^{\frac{1}{2}} (e^{\sigma^2} + 2).$$

Après une transformation logarithmique, on obtient des variables aléatoires normalement distribuées de moyenne  $\mu$  et de variance  $\sigma^2$  données par les formules ci-dessous:

$$\mu = \log(\xi) - \frac{1}{2} \log\left(1 + \frac{\tau}{\exp(2 \log(\xi))}\right),$$

$$\sigma^2 = \sqrt{\log\left(1 + \frac{\tau}{\exp(2 \log(\xi))}\right)} \quad (8)$$

### 2.1.2 Modèle Log-normal Non-stationnaire

Les modèles fréquentiels classiques peuvent être généralisés pour intégrer la notion de non-stationnarité dans le processus de modélisation en intégrant le temps ou d'autres covariables au niveau des paramètres. De telles covariables pourraient incorporer des tendances, des cycles, ou des indices climatiques tels que les indices d'oscillations climatiques de basse fréquence (Coles 2001, Katz *et al.* 2002). Par conséquent ces modèles peuvent être appliqués pour étudier l'effet des changements climatiques, ainsi que l'effet de la variabilité climatique sur la fréquence et l'amplitude des événements extrêmes.

Le modèle Log-normal non-stationnaire correspond au processus  $X_t$ , tel que les paramètres sont fonction du temps ou d'autres covariables. Ainsi, pour des valeurs fixes du temps ou/et de covariables,  $X_t$  est distribué selon une loi Log-normale. Différents types de dépendances peuvent être considérés en représentant des paramètres comme fonctions des covariables. Cependant,

quand deux modèles donnent un bon ajustement, le modèle avec le nombre de paramètres le plus faible sera préféré.

L'introduction des covariables peut être effectuée au niveau de n'importe quel paramètre, ou même à deux ou trois paramètres à la fois. Sankarasubramanian et Lall (2003) ont comparé trois méthodes pour l'estimation des quantiles conditionnels. La première méthode est paramétrique et est basée sur une transformation d'une loi Normale dont les paramètres dépendent de covariables. Les deux autres méthodes correspondent à une approche non paramétrique (Régression des quantiles, Koenkar et Bassett, 1978) et une approche semi-paramétrique (Modèle de vraisemblance locale, Davison et Ramesh, 2000).

L'introduction de covariables au niveau des paramètres implique des changements des caractéristiques statistiques de la distribution ajustée. Plusieurs auteurs suggèrent d'associer un changement dans la variance à celui de la moyenne pour que le coefficient de variation ( $Cv$ ) reste constant. Cependant, dans le cas d'une variable distribuée suivant une loi Log-normale, le coefficient  $Cv$  est fonction seulement du paramètre  $\sigma$ . Ainsi, garder le coefficient  $Cv$  constant implique que le paramètre  $\sigma$  est constant. Par conséquent, dans le présent travail seul le paramètre  $\mu$  est fonction des covariables.

Pour étudier les propriétés des estimateurs des paramètres, trois modèles sont considérés dans le présent travail :

$LN_{(0)}(\mu, \sigma)$ . qui est un modèle classique où tous les paramètres sont constants  $\mu_t = \mu, \sigma_t = \sigma$ .

$LN_{(1)}(\mu_t = \beta_1 + \beta_2 Y_t, \sigma)$ . Le paramètre  $\mu$  est une fonction linéaire d'une covariable  $Y_t$ .

$LN_{(2)}(\mu_t = \beta_1 + \beta_2 Y_t + \beta_3 Y_t^2, \sigma)$ . Le paramètre  $\mu$  est une fonction quadratique d'une covariable temporelle  $Y_t$  et l'autre paramètre est constant.

Notons que pour la distribution Log-normale, le changement du paramètre  $\mu$  mène à un changement de la moyenne et de la variance de la variable, ce qui correspond souvent aux changements observés dans des séries hydrométéorologiques.

## 2.2 Estimation des paramètres par la méthode du maximum de vraisemblance

Les paramètres du modèle Log-normal non-stationnaire peuvent être estimés par la méthode du maximum de vraisemblance, par maximisation du logarithme de la fonction de vraisemblance. La fonction de vraisemblance s'écrit:

$$\begin{aligned}
 L_n(\underline{x} | \mu_t, \sigma_t) &= \prod_{t=1}^n f(x_t; \mu_t, \sigma_t) \\
 &= \prod_{t=1}^n \frac{1}{\sigma_t x_t \sqrt{2\pi}} \exp\left(-\frac{1}{2} \left(\frac{\log x_t - \mu_t}{\sigma_t}\right)^2\right) \\
 &= \frac{1}{\prod_{t=1}^n \sigma_t \prod_{t=1}^n x_t (2\pi)^{n/2}} \prod_{t=1}^n \exp\left(-\frac{1}{2} \left(\frac{\log x_t - \mu_t}{\sigma_t}\right)^2\right);
 \end{aligned} \tag{9}$$

où  $n$  est la taille de la série à ajuster.

Posons  $\sigma_t = \sigma$ , puisque la non-stationnarité considérée dans le présent travail est liée seulement au paramètre  $\mu$ . Dans ce cas la fonction de log-vraisemblance  $l_n$  est donnée par :

$$l_n(\underline{x}; \mu_t, \sigma) = (-n \log(\sigma) - \sum_{t=1}^n \log(x_t) - \frac{n}{2} \log(2\pi)) + \sum_{t=1}^n \left(-\frac{1}{2} \left(\frac{\log x_t - \mu_t}{\sigma}\right)^2\right) \tag{10}$$

Par définition, les estimateurs du maximum de vraisemblance sont les solutions du système d'équations des dérivées partielles de  $l_n$  par rapport à chacun des paramètres.

Pour le modèle  $LN_{(1)}$ , où  $\mu_t = \beta_1 + \beta_2 Y_t$  et  $Y_t$  est le vecteur des covariables, les estimateurs du maximum de vraisemblance des paramètres  $\beta_1, \beta_2$  et  $\sigma$  sont solutions du système d'équations suivant:

$$\left\{ \begin{array}{l} \frac{1}{\sigma^2} \sum_{t=1}^n (\log x_t - \beta_1 - \beta_2 y_t) = 0 \\ \frac{1}{\sigma^2} \sum_{t=1}^n y_t (\log x_t - \beta_1 - \beta_2 y_t) = 0 \\ -\frac{n}{\sigma} + \left( \sum_{t=1}^n \frac{(\log x_t - \beta_1 - \beta_2 y_t)^2}{\sigma^3} \right) = 0 \end{array} \right. \quad (11)$$

Si on pose

$$\left\{ \begin{array}{l} k_1 = \sum_{t=1}^n \log x_t \\ k_2 = \sum_{t=1}^n y_t \\ k_3 = \sum_{t=1}^n y_t \log x_t \\ k_4 = \sum_{t=1}^n y_t^2 \\ k_5 = \sum_{t=1}^n (\log x_t)^2 \end{array} \right. \quad (12)$$

Le système d'équations du maximum de vraisemblance est :

$$\begin{cases} \frac{1}{\sigma^2}(k_1 - n\beta_1 - k_2\beta_2) = 0 \\ \frac{1}{\sigma^2}(k_3 - k_2\beta_1 - k_4\beta_2) = 0 \\ -\frac{n}{\sigma} + \frac{1}{\sigma^3}(k_5 + n\beta_1^2 + k_4\beta_2^2 + 2k_2\beta_1\beta_2 - 2k_1\beta_1 - 2k_3\beta_2) = 0 \end{cases} \quad (13)$$

Ce système peut être résolu directement par substitution ou à partir d'une écriture matricielle équivalente.

D'une manière analogue on obtient un système équivalent au système précédent pour le modèle  $LN_{(2)}$ , avec une quatrième équation pour le paramètre  $\beta_3$ . Les équations correspondant à l'estimation des paramètres du modèle  $LN_{(2)}$  par la méthode du maximum de vraisemblance sont présentées dans l'annexe A.

### 2.3 Estimation des quantiles

Le quantile de probabilité au non-dépassement  $p$  conditionnellement à une valeur observée  $y_0$  de la covariable  $Y$ , est obtenu à partir d'une loi Log-normale de paramètres  $(\mu_{y_0}, \sigma)$ . Ce quantile peut être présenté sous forme explicite :

$$x_{p,y_0} = \exp(\mu_{y_0} + \sigma \Phi^{-1}(p)). \quad (14)$$

Où  $\mu_{y_0}$  est le premier paramètre du modèle Log-normal non-stationnaire conditionnellement à la valeur particulière  $y_0$  de la covariable. Ce paramètre change selon les différents modèles, pour le modèle  $LN_{(0)}$ ,  $\mu_{y_0} = \mu$ , pour le modèle  $LN_{(1)}$  il est fonction linéaire de la covariable:  $\mu_{y_0} = \beta_1 + \beta_2 y_0$ , et pour le modèle  $LN_{(2)}$  il s'écrit sous la forme :

$$\mu_{y_0} = \beta_1 + \beta_2 y_0 + \beta_3 y_0^2.$$

En remplaçant ces paramètres par leurs estimateurs dans l'équation (14), on obtient les estimateurs des quantiles.

### **3 COMPARAISON DES MODÈLES LN ET GEV NON-STATIONNAIRES**

L'objectif de cette section est de comparer le modèle LN non-stationnaire au modèle GEV non-stationnaire. Les modèles GEV non-stationnaires considérés dans ce travail ont été décrits en détail par El Adlouni *et al.* (2007) et sont présentés dans l'annexe B. Pour étudier la différence entre les deux modèles, nous avons considéré la même méthode d'estimation à savoir la méthode du maximum de vraisemblance. Nous avons considéré plusieurs types de non-stationnarité qui peuvent être observées dans des séries hydrométéorologiques. Dans tous les cas, la covariable est distribuée suivant une loi Normale centrée réduite.

#### **3.1 Plan de simulation**

Dans cette partie on compare par simulation les modèles LN et GEV non-stationnaires (Annexe B) dans l'estimation des extrêmes conditionnellement aux valeurs de covariables. Le choix des paramètres des modèles simulés est effectué sur la base de la moyenne et du coefficient de variation de 40 séries de débit maximum annuel du Québec, étudiées par Kouider *et al.* (2002). Les résultats obtenus à partir de cette étude nous ont permis d'obtenir les intervalles de variation des deux paramètres de la loi LN. Les coefficients de variation pour ces 40 séries appartiennent à l'intervalle [0.15; 0.35]. Nous avons considéré pour cette étude deux coefficients de variation :  $C_v = 0.2$  et  $C_v = 0.3$ . La moyenne a été fixée à  $E[X] = 2000$ . Les valeurs de l'asymétrie, pour

une variable distribuée suivant une loi LN, peuvent être déduites de celles du coefficient de variation à partir de la formule suivante :

$$C_s = C_v(C_v^2 + 3) \quad (15)$$

Les coefficients d'asymétrie correspondant aux deux coefficients de variation  $C_v = 0.2$  et  $C_v = 0.3$  sont  $C_s = 0.6$  et  $C_s = 0.9$ , respectivement. Comme conséquence au choix de ces deux cas, les valeurs du paramètre  $\sigma$  pour la loi LN et le paramètre de forme,  $\kappa$ , pour la loi GEV, sont fixés pour chaque asymétrie (Tableau 1).

Comme nous l'avons déjà mentionné, le modèle  $LN_{(1)}$  peut représenter (1) le cas où la moyenne dépend d'une covariable et (2) la moyenne et la variance sont fonction de la covariable. Nous avons considéré les trois cas suivants :

Cas 1 : on compare  $M1 = LN_{(1)}$  et  $M2 = GEV_{(10)}$  (le paramètre de position est fonction linéaire de la covariable).

Cas 2 : on compare  $M1 = LN_{(1)}$  et  $M2 = GEV_{(11)}$  (les paramètres de position et d'échelle sont fonctions linéaires de la covariable).

Cas 3 : comparaison des modèles  $M1 = LN_{(2)}$  et  $M2 = GEV_{(21)}$  (le paramètre de position est une fonction quadratique et le paramètre d'échelle est fonction linéaire de la covariable).

Pour les trois cas, la covariable  $Y_i$  est distribuée selon une loi normale centrée réduite (de moyenne 0 et de variance 1). Les paramètres de dépendance de la covariable, pour les deux modèles, ont été choisis convenablement pour que les échantillons générés à partir des deux modèles aient les mêmes caractéristiques de tendance et de changement d'échelle. Le tableau 1

présente les paramètres des modèles pour les trois cas considérés et pour les deux coefficients de variation.

**[Tableau 1]**

Pour comparer ces modèles (deux à deux), on génère  $N = 5000$  échantillons de taille  $n = 50$  à partir du modèle  $M_i$  ( $i = 1, 2$ ) et on estime les quantiles de probabilité au non-dépassement 50%, 90% et 99% à partir des deux modèles M1 et M2. Les quantiles sont estimés conditionnellement à des valeurs particulières de la covariable (les valeurs : minimale (min), moyenne (moy) et maximale (max)). On calcule ensuite le biais relatif (BR) et la racine de l'erreur quadratique moyenne relative (REQMR). Le BR et la REQMR sont donnés par les formules suivantes :

$$BR = \frac{1}{M} \sum_{i=1}^M \frac{|\hat{Q}_{i,p} - Q_p|}{Q_p} \quad (16)$$

$$REQMR = \left[ \frac{1}{M} \sum_{i=1}^M \left( \frac{\hat{Q}_{i,p} - Q_p}{Q_p} \right)^2 \right]^{1/2} \quad (17)$$

Où  $\hat{Q}_{i,p}$  et  $Q_p$  sont respectivement les estimateurs des quantiles et les quantiles théoriques de probabilité au non-dépassement  $p$  et  $M$  est le nombre de simulations.

Cette comparaison permettra d'évaluer : (1) les erreurs liées à la méthode d'estimation des paramètres pour une faible taille de l'échantillon ( $n = 50$ ) dans le cas où on effectue un bon choix du modèle ainsi que (2) les erreurs liées au mauvais choix du modèle (choisir un modèle LN au lieu de GEV et vice-versa).

### 3.2 Résultats

Les résultats de la comparaison entre les modèles LN et GEV non-stationnaires, pour les 6 cas considérés sont présentés dans les tableaux 2 à 7. Dans chaque cas on présente le BR et le REQMR des quantiles conditionnellement à la valeur minimale, moyenne et maximale de la covariable (distribuée suivant une loi Normale). Notons que la différence entre les deux modèles est relativement importante dans le cas des quantiles extrêmes (probabilité au non-dépassement 99%) conditionnellement à la valeur minimale ou maximale de la covariable. Cependant, l'analyse des résultats va être faite principalement pour le quantile de probabilité au non-dépassement 99% conditionnellement à la valeur maximale de la covariable.

#### Les modèles $LN_{(1)}$ et $GEV_{(10)}$

Pour le premier cas, on compare les modèles  $LN_{(1)}$  et  $GEV_{(10)}$  avec une faible asymétrie (le coefficient de variation  $C_v = 0.2$ ). Les résultats sont présentés dans le tableau 2.

#### [Tableau 2]

On remarque que lorsqu'on génère à partir du modèle  $GEV_{(10)}$  (Bloc de ligne correspondant au modèle  $GEV_{(10)}$ ) le BR du quantile (de probabilité au non-dépassement 99% conditionnellement à la valeur maximale de la covariable) estimé à partir du modèle  $LN_{(1)}$  est de l'ordre de -9% alors qu'il est très faible pour  $GEV_{(10)}$  (-0.75%). Cependant, la différence en terme de la REQMR des quantiles estimés à partir de ces deux modèles, n'est pas aussi importante que celle obtenue pour

le biais. La REQMR est de l'ordre de 12% pour le modèle  $GEV_{(10)}$  et de 14% pour le modèle  $LN_{(1)}$  (pour le même quantile, 99%, et la même valeur de la covariable qui est le maximum).

Dans le cas où les échantillons sont générés à partir d'un modèle  $LN_{(1)}$ , (Bloc de ligne correspondant au modèle  $LN_{(1)}$ ) le BR maximum est de -0.17% pour le modèle  $LN_{(1)}$  et 21% pour le modèle  $GEV_{(10)}$ . On remarque que cette différence est très importante pour le biais et elle l'est aussi pour l'erreur quadratique moyenne (REQMR est de l'ordre de 10% pour  $LN_{(1)}$  et 32% pour  $GEV_{(10)}$ ).

Le tableau 3 présente les résultats de la comparaison de ces deux modèles,  $LN_{(1)}$  et  $GEV_{(10)}$ , pour le coefficient de variation  $C_v = 0.3$ .

### [Tableau 3]

Les mêmes conclusions sont valables pour le deuxième cas ( $C_v = 0.3$ ), avec des valeurs du BR et REQMR relativement plus élevées que celles obtenus dans le cas de faible asymétrie (Tableau 3). Notons que les REQMR des quantiles estimés à partir du modèle  $LN_{(1)}$  sont presque les mêmes quel que soit la population mère (générer à partir du modèle  $GEV_{(10)}$  ou  $LN_{(1)}$ ). Alors que, lorsque l'ajustement est effectué avec le modèle  $GEV_{(10)}$  la différence est très grande dépendamment de la population mère (12% quand on génère à partir de  $GEV_{(10)}$  et 32% dans le cas de valeurs simulées à partir de  $LN_{(1)}$ ).

### Les modèles $LN_{(1)}$ et $GEV_{(11)}$

Pour le deuxième cas, on considère des échantillons générés à partir des modèles  $LN_{(1)}$  et  $GEV_{(11)}$  pour tenir compte de l'effet d'une dépendance linéaire du paramètre d'échelle de la loi GEV en fonction de la covariable. Les résultats dans le cas de  $C_v = 0.2$  sont présentés dans le tableau 4.

#### [Tableau 4]

On remarque que dans ce cas les BR des quantiles estimés à partir du modèle  $GEV_{(11)}$ , pour des échantillons générés à partir du modèle  $LN_{(1)}$ , sont relativement faibles par rapport aux résultats de comparaison des modèles  $GEV_{(10)}$  et  $LN_{(1)}$ . Les BR obtenus sont du même ordre que ceux obtenus avec le modèle  $LN_{(1)}$  lorsque les échantillons sont générés à partir du modèle  $GEV_{(11)}$ . Notons que ces deux modèles se ressemblent plus que ceux considérés dans la première comparaison  $GEV_{(10)}$  et  $LN_{(1)}$ . On remarque également que :

les REQMR des quantiles estimés par le modèle  $GEV_{(11)}$  sont approximativement les mêmes quelle que soit la population à partir de laquelle on génère ;

- les REQMR des quantiles obtenus à partir du modèle  $LN_{(1)}$  sont faibles par rapport à ceux du modèle  $GEV_{(11)}$  quel que soit le modèle à partir duquel on génère. Ceci est dû principalement au nombre de paramètres considérés dans chaque modèle (3 paramètres pour  $LN_{(1)}$  et 5 paramètres pour  $GEV_{(11)}$ ). En effet, les modèles avec plus de paramètres permettent d'avoir une bonne estimation au niveau du biais (relativement faible par

rapport au biais de modèles avec moins de paramètres) alors que l'erreur d'estimation augmente avec le nombre de paramètres.

Les résultats correspondants à la comparaison de ces deux modèles dans le cas de  $C_v = 0.3$  sont donnés dans le tableau 5.

#### **[Tableau 5]**

Les conclusions tirées pour le premier cas, faible asymétrie, restent valables pour  $C_v = 0.3$  avec une légère augmentation du REQMR pour les deux modèles (tableau 5). Cependant, les REQMR des quantiles sont presque les mêmes si on considère le même modèle d'ajustement (càd indépendamment du modèle de génération). Les REQMR des quantiles obtenus par le modèle  $LN_{(1)}$  restent toujours inférieurs à ceux du modèle  $GEV_{(11)}$  quelle que soit la population mère (le modèle considéré pour générer les données,  $GEV_{(11)}$  ou  $LN_{(1)}$ ).

#### **Les modèles $LN_{(2)}$ et $GEV_{(21)}$**

Pour ce troisième cas on compare les modèles  $LN_{(2)}$  et  $GEV_{(21)}$  où les paramètres de position et d'échelle sont, respectivement, des fonctions quadratique et linéaire de la covariable. Ces modèles permettent de modéliser la dépendance des covariables avec plus de flexibilité. Les résultats de cette comparaison sont présentés dans le tableau 6 (cas où  $C_v = 0.2$ ) et le tableau 7 (cas où  $C_v = 0.3$ ).

#### **[Tableau 6]**

On remarque que les REQMR des deux modèles restent les mêmes quelle que soit la distribution de la population mère. Ceci peut être expliqué par la flexibilité de ces deux modèles pour représenter les deux populations considérés dans cette étude (générée à partir du modèle  $GEV_{(21)}$  et du modèle  $LN_{(2)}$ ).

Quand on génère à partir du modèle  $GEV_{(21)}$ , le modèle  $LN_{(2)}$  performe mieux que le modèle  $GEV_{(21)}$  en terme de REQMR, pour les deux asymétries. Ceci est dû principalement à la variance élevée causée par le nombre de paramètre qui est égale à 6 dans le cas du modèle  $GEV_{(21)}$ . Alors que pour le biais, le bon choix du modèle mène automatiquement à un faible biais.

[Tableau 7]

#### **4 APPLICATION : PRÉCIPITATIONS À LA STATION TEHACHAPI**

Nous avons appliqué les modèles GEV et LN non-stationnaires à une série  $\underline{X} = (x_1, \dots, x_n)$  de précipitations maximales annuelles journalières en mm, enregistrées à la station Tehachapi de la Californie (Station USGS #048826 Latitude : 35.13 Longitude : -118.45 Période 1952-2000 et le nombre d'années d'enregistrement  $n = 49$ ). La figure 1 montre la localisation de la station d'étude. Située au sud de la Californie, les précipitations de cette région sont fortement modulées par le phénomène El Niño.

[Figure 1]

L'objectif de cette application est d'illustrer la possibilité de tenir compte dans la distribution des précipitations maximales annuelles, des dépendances de certains indices climatiques. Pour la présente étude on considère l'indice SOI « Southern Oscillation Index ». On ajuste le modèle Log-normal non-stationnaire sous ces deux formes  $LN_{(1)}$  et  $LN_{(2)}$  pour analyser la dépendance des paramètres de l'indice SOI. Le modèle GEV non-stationnaire sera également considéré pour montrer la différence entre les deux familles de modèles. Le coefficient de corrélation entre la variable  $X$  et la covariable SOI est -0.57. La figure 2 illustre la relation entre ces deux variables.

**[Figure 2]**

La figure 3 montre qu'il y a une corrélation négative significative (On obtient un coefficient de corrélation  $\rho = -0.57$ ) entre les précipitations et l'indice climatique SOI. Le test de Chi-deux montre que cette corrélation est significative. On remarque également que les plus grandes valeurs enregistrées pour les précipitations maximales annuelles correspondent aux faibles valeurs de l'indice climatique SOI.

**[Figure 3]**

Tous les modèles, LN et GEV, ont été ajustés à la série des précipitations maximales annuelles en considérant comme covariable SOI. On s'intéresse principalement à l'estimation de la médiane (quantile correspondant à la probabilité au non-dépassement 50%) par les trois modèles dans le

cas de LN ( $LN_{(0)}$ ,  $LN_{(1)}$  et  $LN_{(2)}$ ) et quatre modèles dans le cas de GEV ( $GEV_{(0)}$ ,  $GEV_{(10)}$ ,  $GEV_{(11)}$  et  $GEV_{(21)}$ ).

Le choix du modèle est fait à l'aide de la déviance statistique. Le test basé sur la déviance est simple et très utilisé pour comparer deux modèles tels que l'un est un cas particulier de l'autre (Coles, 2001). Ce test est utilisé, séparément, pour choisir un modèle parmi les modèles LN non-stationnaires et un parmi les modèles GEV non-stationnaires. Quand la différence entre les deux modèles n'est pas évidente, il est préférable d'utiliser le modèle le plus simple afin de respecter le principe de parcimonie.

Pour deux modèles  $M_1$  et  $M_0$  tels que  $M_0 \subset M_1$ , la statistique de déviance est définie par:  $D = 2\{l_n^*(M_1) - l_n^*(M_0)\}$ , où  $l_n^*(M)$  est la valeur optimale du log-vraisemblance du modèle  $M$ . Une grande valeur de  $D$  indique que le modèle  $M_1$  est le plus adéquat que le modèle  $M_0$  pour ajuster les variations des données. La statistique  $D$  est distribuée selon une loi de chi-deux ( $\chi_v^2$ ). Le paramètre  $v$  est égal à la différence entre la dimension (le nombre de paramètres) du modèle  $M_1$  et la dimension du modèle  $M_0$ . Des valeurs de  $D$  qui sont plus grandes que les quantiles de la loi de  $\chi_v^2$  pour un niveau de confiance particulier  $\alpha = 5\%$ , signifie que le modèle  $M_1$  est meilleur que le modèle  $M_0$ .

Les résultats montrent que pour les modèles LN non-stationnaires, il y a une différence significative entre les modèles  $LN_{(1)}$  et  $LN_{(0)}$  puisque  $D = 15.49$ . Cette valeur est plus grande que le quantile (95%) de la loi  $\chi_1^2$  ( $\Pr(\chi_1^2 \leq 15.49) = 0.9999$ ). D'autre part, pour les modèles  $LN_{(1)}$  et  $LN_{(2)}$ , le test statistique montre qu'il n'y a pas de différence significative entre ces deux

modèles, car  $D = 1.20$  et  $\Pr(\chi_1^2 \leq 1.20) = 0.7273$ . Ainsi, pour le niveau de signification  $\alpha = 5\%$ , le modèle  $LN_{(1)}$  est le plus adéquat pour représenter la dépendance entre les précipitations maximales annuelles et la covariable SOI. Les estimations des paramètres ainsi que la valeur de la fonction log-vraisemblance correspondant à chacun des modèles LN, sont données dans le tableau 8.

**[Tableau 8]**

Le test basé sur la statistique de déviance a été également appliqué aux modèles GEV non-stationnaires. Les résultats montrent que le modèle  $GEV_{(21)}$  est le plus adéquat pour représenter la dépendance entre les précipitations maximales annuelles et l'indice SOI. Le tableau 9 présente les estimations des paramètres pour chaque modèle ainsi que les valeurs de la fonction de log-vraisemblance considérées pour le test de la déviance.

**[Tableau 9]**

**[Figure 4]**

Les médianes sont calculées conditionnellement à plusieurs valeurs de la covariable pour l'estimation de la courbe médiane. Les résultats correspondant à certaines valeurs particulières (la valeur minimale, la moyenne et la valeur maximum) sont donnés dans le tableau 10. Les estimateurs de la médiane obtenus par les deux modèles retenus,  $GEV_{(21)}$  et  $LN_{(1)}$ , sont

représentés dans la figure 4. Les intervalles de confiance sont calculés en générant  $N = 5000$  échantillons par Bootstrap. On remarque qu'il y a une légère différence entre les médianes conditionnelles obtenues à partir des deux modèles, surtout au niveau des valeurs minimale et maximale.

### [Tableau 10]

Dans le cas stationnaire pour les modèles  $LN_{(0)}$  et  $GEV_{(0)}$ , les estimations de la médiane sont respectivement égale à 27.12 et 24.96. Ces valeurs correspondent à la médiane estimée par les modèles non-stationnaires conditionnellement à la valeur moyenne de la covariable. La différence entre les modèles stationnaires et non-stationnaires est très grande au niveau des valeurs extrêmes de la covariable (surtout pour la valeur minimale). Quel que soit le modèle non-stationnaire choisi, LN ou GEV, l'estimation de la médiane peut passer du simple au double (même plus que le double pour quelques valeurs de la covariable) dépendamment du fait qu'on tienne compte de la dépendance de la covariable ou non. Pour les grandes valeurs de la covariable SOI, les estimations conditionnelles de la médiane sont inférieures à celles obtenues dans le cas stationnaire.

Notons que le modèle  $GEV_{(21)}$  est tel que le paramètre de position est une fonction quadratique de la covariable. Le modèle LN non-stationnaire qui ressemble le plus au modèle  $GEV_{(21)}$  est le modèle  $LN_{(2)}$ . Ce modèle n'a pas été retenu par le test basé sur la statistique de la déviance, puisque la différence avec le modèle  $LN_{(1)}$  n'est pas significative. La figure 5 présente les

médianes conditionnelles obtenues à partir des modèles  $GEV_{(21)}$  et  $LN_{(2)}$ . On remarque que ces deux modèles donnent des résultats très semblables.

[Figure 5]

## 5. CONCLUSIONS

L'objectif principal de cette étude était de présenter le modèle Log-normal non-stationnaire pour l'ajustement des séries avec tendances ou qui dépendent de covariables. Cette dépendance est introduite au niveau des paramètres pas le biais de fonctions usuelles et des hyper-paramètres. Le modèle GEV est le modèle le plus étudié pour représenter ces dépendances. L'introduction du modèle Log-normal non-stationnaire et sa comparaison avec le modèle GEV non-stationnaire nous a permis d'étudier la flexibilité du modèle LN pour représenter ces dépendances et d'évaluer les erreurs liées à la méthode d'estimation ou celles générées par un mauvais choix du modèle. La loi GEV est une loi à trois paramètres, elle est, par conséquent, une des lois les plus flexibles pour l'ajustement des extrêmes. La loi LN est également très utilisée pour l'analyse des extrêmes vu qu'elle possède des propriétés similaires aux distributions des extrêmes (au niveau des queues des distributions).

Les résultats de simulation montrent que les modèles LN non-stationnaires performant bien même dans le cas de séries générées à partir d'un modèle GEV non-stationnaire, en terme d'erreur moyenne quadratique relative. En effet, dans tous les cas considérés, les REQMR des quantiles estimés par les modèles LN non-stationnaires, sont plus petits que ceux obtenus à partir des modèles GEV non-stationnaires. Le biais est souvent lié au bon choix de la distribution. On

en déduit que: (1) la flexibilité des modèles LN non-stationnaires pour représenter ce type de dépendance et (2) la nécessité de développer de nouveaux critères, autres que la statistique de déviance, pour le choix de l'ajustement le plus adéquat dans le cas des modèles non-stationnaires.

L'application des deux familles de modèles à des données de précipitations maximales annuelles conditionnellement à l'indice climatique SOI, illustre la flexibilité de ces modèles pour l'ajustement des séries hydrométéorologiques. Le cas étudié, montre également, l'avantage des modèles non-stationnaires pour incorporer l'information additionnelle. En fait, pour les petites valeurs de la covariable SOI, le modèle classique sous-estime la médiane ce qui peut avoir des conséquences très dramatiques. L'utilisation des estimations conditionnelles a permis de déduire les valeurs des quantiles correspondants aux événements particuliers représentés par la covariable.



## RÉFÉRENCES

- Chen H. et Rao A. (2002).** Testing Hydrologic Time Series for Stationarity, Newspaper of Hydrologic Engineering, vlo. 7, no. (2). pp. 129-136.
- Clarke R.T. (2002a).** Estimating time trends in Gumbel-distributed data by means of generalized linear models, Water Ressources Research, vol. 38, no.(7)., 1111, 10.1029/2001 WR000917, pp. 16-1/16-11.
- Clarke, R. T. (2002b).** Estimating trends in data from the Weibull and a generalized extreme value distribution, Water Resources Research, 38(6), 1089, doi:10.1029/2001WR000575.
- Coles G.S. (2001).** An Introduction to Statistical Modeling of Extreme Value, Springer, 208 p.
- Davison, A. C. et Ramesh, N. I. (2000).** Local likelihood smoothing of sample extremes. Journal of the Royal Statistical Society Series B, Statistics in Society, Volume 62 part 1, pages 191-208.
- El Adlouni S., Ouarda T. BMJ., Zhang X., Roy R. et Bobée B. (2007).** Generalized maximum likelihood estimators of the non-stationary GEV model parameters. Water Resources Research. Sous presse.
- Faucher, D., Ouarda, T.B.M.J. et Bobée B. (1997).** Revue bibliographique des tests de stationnarité. INRS-Eau, rapport de Recherche no R-499, 66 pages.
- Hobijn B. (2004).** Generalizations of the KPSS-test for Stationarity, Statistica Neerlandica, vol. 58, no. (4). pp. 483 - 502.
- Hosking, J.R.M. (1990).** L-Moments: Analysis and Estimation of Distributions Using Linear Combinations of Order Statistics. Newspaper of Royal Statistical Society, 52:105-124.
- Katz R.W., Parlange M.B. et Naveau P. (2002).** Statistics of extremes in hydrology, Advances in Water Resources, vol. 25, pp. 1287-1304.
- Kim K.-D. et Heo J.-H (2002).** Comparative study of flood quantiles estimation by non parametric models. Journal of Hydrology. vol 260, pp. 176-193.

- Koenkar, R. et Bassett G. S. (1978).** Regression quantiles, *Econometrica*, 46, 33–50, 1978.
- Kouider, A. Gingras, H., Ouarda, T. B.M.J., Ristic-Rudolf, Z. et Bobée, B. (2002).** Analyse fréquentielle locale et régionale et cartographie des crues au Québec, Rap. *R-627-el*, INRS-ETE, Ste-Foy, Canada.
- Leurent F. (1998).** Les valeurs du temps des automobilistes à Marseille en 1995: The distribution of car drivers' values-of-time in Marseilles (1995), *Recherche - Transports – Sécurité*, Vol.60, pp.19-38.
- Martin E.S et Stedinger J.R. (2000).** Generalized Maximum Likelihood GEV Quantile Estimators for Hydrologic Data. *Water Resources Research*, vol. 36. No. (3). pp.737-744.
- Önöz B. et Bayazit M (2003).** The Power of Statistical Tests for Trend Detectio. *Turkish J. Eng. Env. Sci.* 27, pp. 247 - 251.
- Ouarda, T. B.M.J., Ashkar, F., Bensaïd, E. et Hourani I. (1994).** Distributions statistiques utilisées en hydrologie: Transformations et propriétés asymptotiques. Département de mathématique, Université de Moncton, 31 p.
- Sankarasubramanian, A. et Lall U. (2003).** Flood Quantiles and Changing Climate: Seasonal forecasts and causal relations. *Water Resources Research*, 39(5):1134.
- USWRC (1982).** Guidelines for Determining Flood Flow Frequential. Bulletin 17B, Hydrology Committee, (U.S. Water Resources Council):Washington, DC.
- Zhang X., K. D. Harvey, W. D. Hogg, et Yuzyk T. R. (2001).** Trends in Canadian streamflow. *Water Resources Research*, Vol. 37, No. 4, pp. 987–999.

## LISTE DE TABLEAUX

Tableau 1: Les paramètres des lois LN et GEV correspondant aux deux coefficients de variation considérés.

Tableau 2: BR et REQMR des quantiles estimés, conditionnellement à la covariable, par le modèle  $GEV_{(10)}$  et  $LN_{(1)}$  pour  $C_v = 0.2$ .

Tableau 3: BR et REQMR des quantiles estimés, conditionnellement à la covariable, par le modèle  $GEV_{(10)}$  et  $LN_{(1)}$  pour  $C_v = 0.3$ .

Tableau 4: BR et REQMR des quantiles estimés, conditionnellement à la covariable, par le modèle  $GEV_{(11)}$  et  $LN_{(1)}$  pour  $C_v = 0.2$ .

Tableau 5: BR et REQMR des quantiles estimés, conditionnellement à la covariable, par le modèle  $GEV_{(11)}$  et  $LN_{(1)}$  pour  $C_v = 0.3$ .

Tableau 6: BR et REQMR des quantiles estimés, conditionnellement à la covariable, par le modèle  $GEV_{(21)}$  et  $LN_{(1)}$  pour  $C_v = 0.2$ .

Tableau 7: BR et REQMR des quantiles estimés, conditionnellement à la covariable, par le modèle  $GEV_{(21)}$  et  $LN_{(2)}$  pour  $C_v = 0.3$ .

Tableau 8: Les estimateurs des paramètres et les valeurs de la fonction log-vraisemblance correspondantes des modèles LN non-stationnaires.

Tableau 9: Les estimateurs des paramètres et les valeurs de la fonction log-vraisemblance correspondantes des modèles GEV non-stationnaires.

Tableau 10: Les estimateurs des quantiles de probabilité au non-dépassement 50% conditionnellement aux valeurs minimale, moyenne et maximale de SOI.

## LISTE DE FIGURES

Figure 1: Situation géographique de la station Tehachapi de la Californie.

Figure 2: Précipitations maximales annuelles observées et les valeurs des SOI correspondantes.

Figure 3: Nuage de points des précipitations maximales annuelles observées et les valeurs des SOI correspondantes.

Figure 4: La médiane Conditionnelle et ses intervalles de confiance estimés par les modèles  $GEV_{(21)}$  et  $LN_{(1)}$ .

Figure 5: La médiane Conditionnelle et ses intervalles de confiance estimés par les modèles  $LN_{(2)}$  et  $GEV_{(21)}$ .

**Tableau 1: Les paramètres des lois LN et GEV correspondant aux deux coefficients de variation considérés.**

$C_v$	$C_v = 0.2$	$C_v = 0.3$
	$\sigma = 0.193$ pour <i>LN</i> & $\kappa = 0.11$ pour <i>GEV</i>	$\sigma = 0.283$ pour <i>LN</i> & $\kappa = 0.04$ pour <i>GEV</i>
<b>Cas 1</b>	$LN_{(1)}(\mu_t = 6.2 - 0.1Y_t, \sigma)$ & $GEV_{(10)}(\mu_t = 1060 - 200Y_t, 200, \kappa)$	$LN_{(1)}(\mu_t = 7.28 - 0.1Y_t, \sigma)$ & $GEV_{(10)}(\mu_t = 1360 - 200Y_t, 250, \kappa)$
<b>Cas 2</b>	$LN_{(1)}(\mu_t = 6.2 - 0.1Y_t, \sigma)$ & $GEV_{(11)}(1060 - 200Y_t, \exp(\log(200) - 0.2Y_t), \kappa)$ $LN_{(2)}(\mu_t = 6.2 - 0.2Y_t + 0.1Y_t^2, \sigma)$	$LN_{(1)}(\mu_t = 7.28 - 0.1Y_t, \sigma)$ & $GEV_{(11)}(1360 - 200Y_t, \exp(\log(250) - 0.2Y_t), \kappa)$ $LN_{(2)}(\mu_t = 7.28 - 0.2Y_t + 0.1Y_t^2, \sigma)$
<b>Cas 3</b>	& $GEV_{(21)}(1060 - 250Y_t + 3Y_t^2, \exp(\log(200) - 0.2Y_t), \kappa)$	& $GEV_{(21)}(1360 - 250Y_t + 6Y_t^2, \exp(\log(250) - 0.2Y_t), \kappa)$

**Tableau 2: BR et REQMR des quantiles estimés, conditionnellement à la covariable, par le modèle  $GEV_{(10)}$  et  $LN_{(1)}$  pour  $C_v = 0.2$ .**

Distribution à générer	Valeurs de la covariable	BR % ( $GEV_{(10)}$ )			BR % ( $LN_{(1)}$ )		
		50%	90%	99%	50%	90%	99%
$GEV_{(10)}$	Min	0,1167	-0,6071	-0,5971	2,3488	4,9048	7,1779
	Moy	0,1792	-0,652	-0,6328	-0,0265	-0,8183	-0,7808
	Max	0,3385	-0,7544	<b>-0,7088</b>	3,5091	-5,0805	<b>-9,3557</b>
$LN_{(1)}$	Min	-7,0558	-12,2032	-15,3113	-0,281	-0,965	-1,4717
	Moy	1,1965	1,161	1,4895	-0,0447	-0,7298	-1,2373
	Max	-14,9357	6,8264	<b>21,6301</b>	1,0235	0,3334	<b>-0,1777</b>
		REQMR % ( $GEV_{(10)}$ )			REQMR % ( $LN_{(1)}$ )		
$GEV_{(10)}$	Min	5,7008	5,2217	8,1413	7,2294	9,0174	11,3401
	Moy	3,2666	3,8618	8,1967	2,9377	3,9296	5,4275
	Max	13,9238	10,6692	<b>12,2872</b>	12,397	12,2267	<b>14,546</b>
$LN_{(1)}$	Min	8,9925	13,0125	16,7522	6,3411	6,8431	7,8802
	Moy	3,3721	4,1555	10,0396	2,6972	3,7603	5,4424
	Max	22,7677	17,2334	<b>32,1192</b>	9,696	9,9727	10,6891

**Tableau 3: BR et REQMR des quantiles estimés, conditionnellement à la covariable, par le modèle  $GEV_{(10)}$  et  $LN_{(1)}$  pour  $C_v = 0.3$ .**

Distribution à générer	Valeurs de la covariable	BR % ( $GEV_{(10)}$ )			BR % ( $LN_{(1)}$ )		
		50%	90%	99%	50%	90%	99%
$GEV_{(10)}$	Min	-0,1135	-0,8206	-0,2441	3,6135	6,4569	6,5908
	Moy	0,014	-0,836	-0,1855	0,3606	-1,3428	-4,1929
	Max	0,3865	-0,8743	<b>-0,054</b>	7,0594	-6,548	<b>-15,5546</b>
$LN_{(1)}$	Min	-8,8234	-15,2802	-17,0661	0,5907	-0,1787	-0,697
	Moy	1,0043	1,2334	4,2694	0,0498	-0,7099	-1,221
	Max	-8,3182	20,2805	<b>42,3209</b>	0,7385	-0,0176	<b>-0,5253</b>
		REQMR % ( $GEV_{(10)}$ )			REQMR % ( $LN_{(1)}$ )		
$GEV_{(10)}$	Min	5,9431	5,6734	10,6852	8,4919	10,9184	12,1733
	Moy	3,2936	4,4303	11,4481	3,0372	4,4936	7,3318
	Max	16,2435	11,8448	16,1874	15,5379	13,895	<b>19,3964</b>
$LN_{(1)}$	Min	11,9007	16,4063	20,3055	9,6373	10,1857	11,6048
	Moy	4,8246	6,0558	16,9621	4,1189	5,5244	7,9274
	Max	25,7549	29,3016	<b>55,3627</b>	14,7056	15,0891	16,102

**Tableau 4: BR et REQMR des quantiles estimés, conditionnellement à la covariable, par le modèle  $GEV_{(1)}$  et  $LN_{(1)}$  pour  $C_v = 0.2$ .**

Distribution à générer	Valeurs de la covariable	BR % ( $GEV_{(1)}$ )			BR % ( $LN_{(1)}$ )		
		50%	90%	99%	50%	90%	99%
$GEV_{(1)}$	Min	0,5326	2,0971	3,9755	1,5384	-3,898	<b>-6,5004</b>
	Moy	-0,1191	-1,2631	-1,3029	-0,2612	-1,0188	<b>-0,9068</b>
	Max	0,5967	-0,0423	0,1261	4,4623	5,0302	<b>7,8067</b>
$LN_{(1)}$	Min	-5,9857	-3,975	<b>-4,0247</b>	0,7624	0,0054	-0,562
	Moy	1,3396	0,0495	<b>-1,8116</b>	0,2707	-0,4743	-1,032
	Max	-6,6920	-6,6778	<b>-7,0340</b>	0,2563	-0,4751	-1,0218
		REQMR % ( $GEV_{(1)}$ )			REQMR % ( $LN_{(1)}$ )		
$GEV_{(1)}$	Min	6,6336	13,024	22,5875	7,3296	8,1469	<b>10,1708</b>
	Moy	3,1937	4,0324	8,7902	2,751	3,9203	<b>5,5304</b>
	Max	12,2919	15,1863	21,382	11,2984	14,5099	<b>18,0267</b>
$LN_{(1)}$	Min	7,8961	9,9393	<b>15,2115</b>	6,5905	6,8570	7,7323
	Moy	3,4299	3,9168	<b>8,9044</b>	2,7551	3,7579	5,3925
	Max	21,5604	17,4905	<b>22,5959</b>	10,3707	10,746	11,4793

**Tableau 5: BR et REQMR des quantiles estimés, conditionnellement à la covariable, par le modèle  $GEV_{(1)}$  et  $LN_{(1)}$  pour  $C_v = 0.3$ .**

Distribution à générer	Valeurs de la covariable	BR % ( $GEV_{(1)}$ )			BR % ( $LN_{(1)}$ )		
		50%	90%	99%	50%	90%	99%
$GEV_{(1)}$	Min	0,4383	0,9757	2,4842	3,9459	7,1784	<b>7,5777</b>
	Moy	0,1391	-1,0062	-0,578	0,4898	-1,22	<b>-4,0858</b>
	Max	1,0246	1,0181	2,6772	6,5788	-7,7658	<b>-17,0544</b>
$LN_{(1)}$	Min	-6,1458	-4,2209	<b>-7,8524</b>	0,8957	0,2674	-0,1458
	Moy	1,0028	-1,3825	<b>-3,4656</b>	0,0900	-0,5274	-0,9328
	Max	-11,7187	-6,705	<b>-15,8418</b>	0,4567	-0,1512	-0,5491
		REQMR % ( $GEV_{(1)}$ )			REQMR % ( $LN_{(1)}$ )		
$GEV_{(1)}$	Min	6,5147	10,3798	18,2413	8,4344	11,1209	<b>12,4168</b>
	Moy	3,3386	4,5496	11,8291	3,0590	4,4717	<b>7,2161</b>
	Max	17,7386	24,004	34,8104	14,7294	14,0394	<b>20,3451</b>
$LN_{(1)}$	Min	9,632	13,4981	<b>22,6546</b>	10,0321	10,5355	11,8139
	Moy	4,7207	5,5427	<b>13,0476</b>	4,0581	5,4298	7,6967
	Max	25,3414	23,0835	<b>31,1814</b>	14,8734	15,3201	16,2949

**Tableau 6: BR et REQMR des quantiles estimés, conditionnellement à la covariable, par le modèle  $GEV_{(21)}$  et  $LN_{(1)}$  pour  $C_v = 0.2$ .**

Distribution à générer	Valeurs de la covariable	BR % (GEV)			BR % (LN)		
		50%	90%	99%	50%	90%	99%
$GEV_{(21)}$	Min	1,5574	3,0286	4,9800	-0,4035	-4,3754	<b>-6,275</b>
	Moy	-0,1724	-1,6317	-1,8156	0,3392	-0,9145	<b>-1,1964</b>
	Max	4,9272	2,0970	0,9122	2,6548	2,9637	<b>3,6896</b>
$LN_{(2)}$	Min	-6,5892	-7,0812	<b>-7,176</b>	1,198	0,2495	-0,471
	Moy	0,5024	0,8044	<b>1,0925</b>	0,1198	-0,8108	-1,5173
	Max	3,4608	-11,6048	<b>-9,4687</b>	1,2994	0,3537	-0,364
		REQMR % (GEV)			REQMR % (LN)		
$GEV_{(21)}$	Min	10,0816	14,1395	22,2985	10,4896	11,2024	<b>12,4056</b>
	Moy	3,5489	4,2215	9,1112	3,2035	4,1673	<b>5,7043</b>
	Max	20,7593	19,7870	24,8508	17,5166	17,8638	<b>18,6001</b>
$LN_{(2)}$	Min	10,3698	12,357	<b>17,381</b>	10,0273	10,1049	10,6776
	Moy	3,6232	4,2909	<b>9,7341</b>	3,106	4,0418	5,6433
	Max	19,5203	19,9147	<b>26,38</b>	18,0044	17,9458	18,2092

**Tableau 7: BR et REQMR des quantiles estimés, conditionnellement à la covariable, par le modèle  $GEV_{(21)}$  et  $LN_{(2)}$  pour  $C_v = 0.3$ .**

Distribution à générer	Valeurs de la covariable	BR % (GEV)			BR % (LN)		
		50%	90%	99%	50%	90%	99%
$GEV_{(21)}$	Min	2,2208	3,4686	5,5788	1,5315	4,0845	3,8927
	Moy	-0,3366	-1,6575	-1,4929	1,0383	-1,3151	-4,7978
	Max	6,7546	3,2818	2,4839	2,5833	-8,872	-17,1624
$LN_{(2)}$	Min	-6,3509	-8,0337	-7,5085	0,7935	-0,4121	-1,2904
	Moy	0,2201	0,4246	2,0317	0,0566	-1,1266	-1,9878
	Max	4,5947	-15,9064	-27,4412	2,2419	1,0192	0,127
		REQMR % (GEV)			REQMR % (LN)		
$GEV_{(21)}$	Min	9,1949	13,3321	21,8922	9,4219	10,82	<b>11,607</b>
	Moy	3,7253	4,8029	10,9503	3,6162	4,8316	<b>7,7923</b>
	Max	29,8777	29,2712	37,5546	26,3618	24,9923	<b>27,4917</b>
$LN_{(2)}$	Min	13,1193	16,1901	<b>24,4044</b>	13,8934	14,0933	14,9762
	Moy	5,3564	6,0903	<b>14,6999</b>	4,5538	5,8822	8,1071
	Max	27,3967	26,8065	<b>34,2027</b>	25,3637	25,2076	25,5278

**Tableau 8: Les estimateurs des paramètres et les valeurs de la fonction log-vraisemblance correspondantes des modèles LN non-stationnaires.**

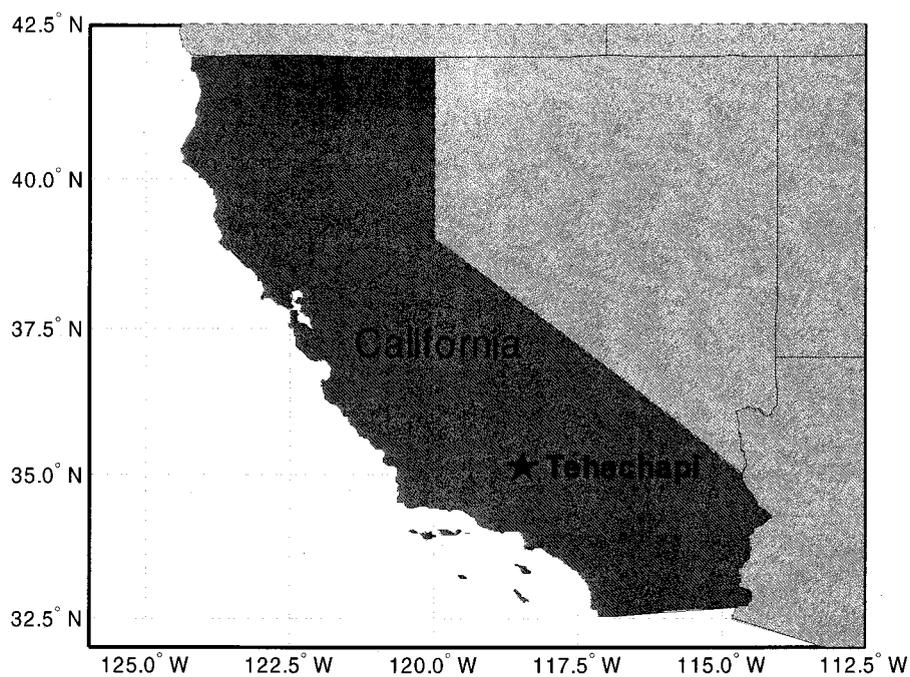
	$l_n^*$	$\mu_1$	$\mu_2$	$\mu_3$	$\sigma$
$LN_{(0)}$	<b>-195.69</b>	3.301	-	-	0.484
$LN_{(1)}$	<b>-187.94</b>	3.256	-0.287	-	0.413
$LN_{(2)}$	<b>-187.34</b>	3.226	-0.259	0.043	0.408

**Tableau 9: Les estimateurs des paramètres et les valeurs de la fonction log-vraisemblance correspondantes des modèles GEV non-stationnaires.**

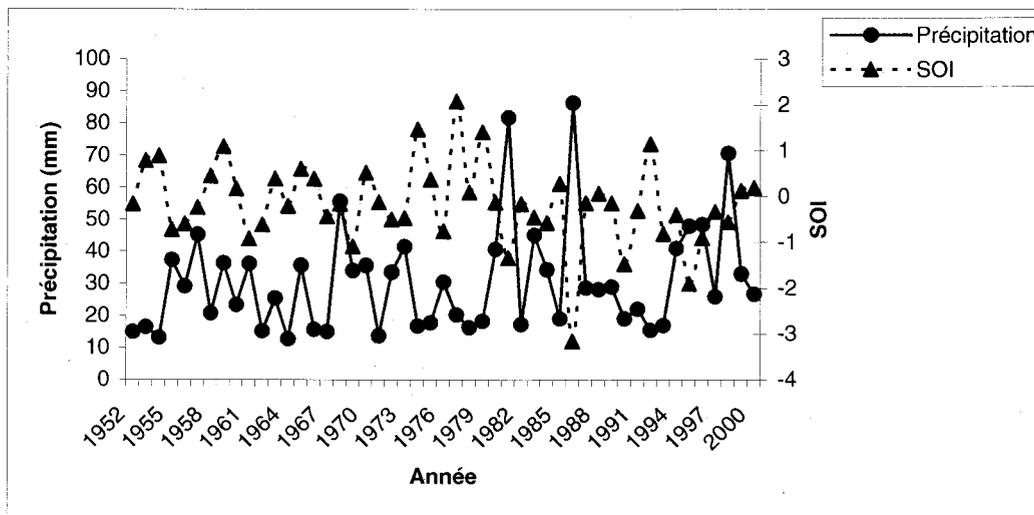
	$l_n^*$	$\mu_1$	$\mu_2$	$\mu_3$	$\alpha_1$	$\alpha_2$	$\kappa$
$GEV_{(0)}$	<b>-194.620</b>	21.503	-	-	2.169	-	-0.308
$GEV_{(10)}$	<b>-193.288</b>	22.530	-3.364	-	2.247	-	-0.208
$GEV_{(11)}$	<b>-187.791</b>	22.482	-3.747	-	2.126	-0.480	-0.075
$GEV_{(21)}$	<b>-185.744</b>	20.469	-6.599	3.192	1.982	-0.672	-0.194

**Tableau 10: Les estimateurs des quantiles de probabilité au non-dépassement 50% conditionnellement aux valeurs minimale, moyenne et maximale de SOI.**

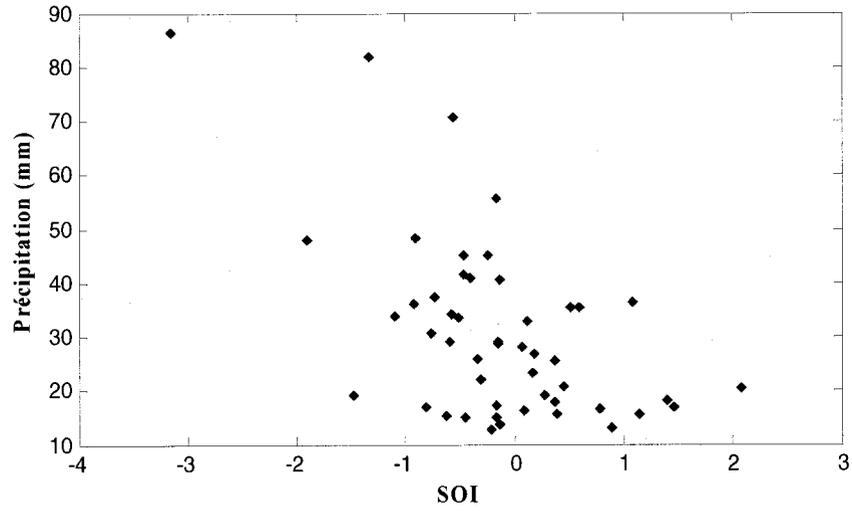
	SOI=-3.1600	SOI=-0.1553	SOI= 2.0720
$LN_{(1)}$	<b>64,2678</b>	<b>27,1290</b>	<b>14,3147</b>
$GEV_{(21)}$	<b>76.3131</b>	<b>24.6317</b>	<b>21.1874</b>



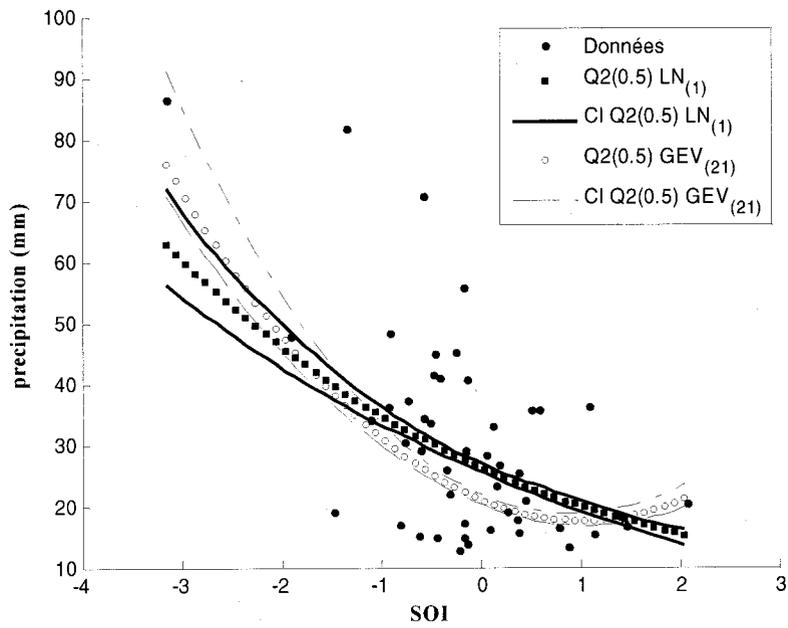
**Figure 1. Situation géographique de la station Tehachapi de la Californie.**



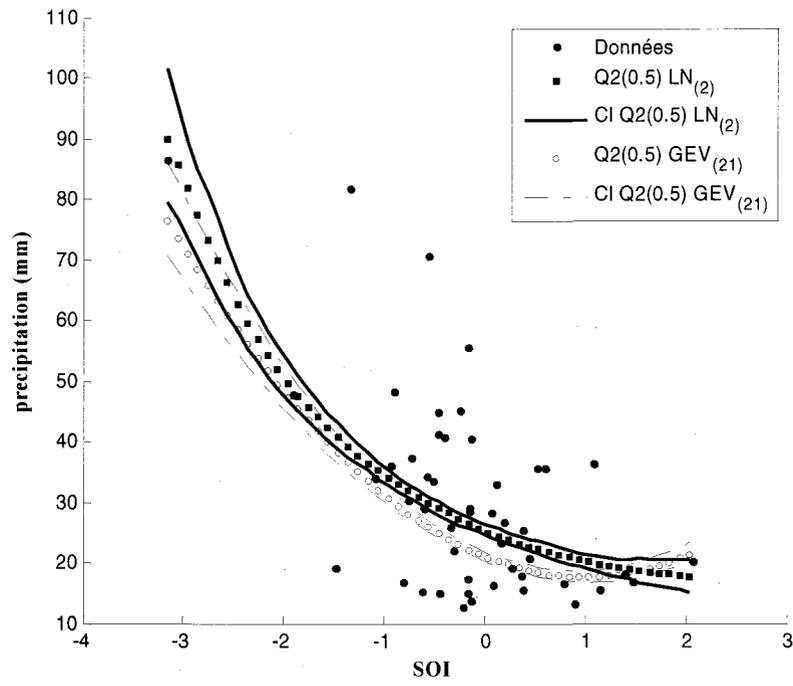
**Figure 2. Précipitations maximales annuelles observées et les valeurs des SOI correspondantes**



**Figure 3. Nuage de points des précipitations maximales annuelles observées et les valeurs des SOI correspondantes.**



**Figure 4. La médiane Conditionnelle et ses intervalles de confiance estimés par les modèles  $GEV_{(21)}$  et  $LN_{(1)}$ .**



**Figure 5.** La médiane Conditionnelle et ses intervalles de confiance estimés par les modèles  $LN_{(2)}$  et  $GEV_{(21)}$ .

## Annexe A

La fonction de log-vraisemblance  $l_n$  pour le modèle LN non-stationnaire est donnée par :

$$l_n(\underline{x}; \mu_t, \sigma) = (-n \log(\sigma) - \sum_{t=1}^n \log(x_t) - \frac{n}{2} \log(2\pi)) + \sum_{t=1}^n \left( -\frac{1}{2} \left( \frac{\log x_t - \mu_t}{\sigma} \right)^2 \right) \quad (A1)$$

Par définition, les estimateurs par maximum de vraisemblance sont les solutions du système d'équation des dérivées partielles de  $l_n$  par rapport à chacun des paramètres.

Pour le modèle  $LN2_{(2)}$  ou on a  $\mu_t = \mu_1 + \mu_2 Y_t + \mu_3 Y_t^2$

où  $Y_t$  est le vecteur des covariables, les estimateurs du maximum de vraisemblance des paramètres

$\mu_1, \mu_2, \mu_3$  et  $\sigma$ , seront solutions du système d'équations suivant:

$$\begin{cases} \frac{1}{\sigma^2} \sum_{t=1}^n (\log x_t - \mu_1 - \mu_2 Y_t - \mu_3 Y_t^2) = 0 \\ \frac{1}{\sigma^2} \sum_{t=1}^n Y_t (\log x_t - \mu_1 - \mu_2 Y_t - \mu_3 Y_t^2) = 0 \\ \frac{1}{\sigma^2} \sum_{t=1}^n Y_t^2 (\log x_t - \mu_1 - \mu_2 Y_t - \mu_3 Y_t^2) = 0 \\ -\frac{n}{\sigma} + \left( \frac{1}{\sigma^3} \sum_{t=1}^n (\log x_t - \mu_1 - \mu_2 Y_t - \mu_3 Y_t^2)^2 \right) \end{cases}$$

Si on pose

$$\left\{ \begin{array}{l} k_1 = \sum_{i=1}^n \log x_i \\ k_2 = \sum_{i=1}^n Y_i \\ k_3 = \sum_{i=1}^n Y_i \log x_i \\ k_4 = \sum_{i=1}^n Y_i^2 \\ k_5 = \sum_{i=1}^n (\log x_i)^2 \\ k_6 = \sum_{i=1}^n Y_i^3 \\ k_7 = \sum_{i=1}^n Y_i^4 \\ k_8 = \sum_{i=1}^n Y_i^2 \log x_i \end{array} \right. \quad (\text{A2})$$

le système d'équations du maximum de vraisemblance est :

$$\left\{ \begin{array}{l} \frac{1}{\sigma^2} (k_1 - n\mu_1 - k_2\mu_2 - k_4\mu_3) = 0 \\ \frac{1}{\sigma^2} (k_3 - k_2\mu_1 - k_4\mu_2 - k_6\mu_3) = 0 \\ \frac{1}{\sigma^2} (k_8 - k_4\mu_1 - k_6\mu_2 - k_7\mu_3) = 0 \\ -\frac{n}{\sigma} + \frac{1}{\sigma^3} (k_5 + n\mu_1^2 + k_4\mu_2^2 + 2k_2\mu_1\mu_2 - 2k_1\mu_1 - 2k_3\mu_2 - 2\mu_3k_8 + 2k_4\mu_1\mu_3 + 2k_6\mu_2\mu_3 + k_7\mu_3^2) = 0 \end{array} \right.$$

Ce système peut être résolu par une méthode numérique.

## Annexe B

### Quelques exemple du modèle GEV non-stationnaire.

1.  $GEV_{(0)}(\mu, \alpha, \kappa)$  : correspond au modèle stationnaire où les paramètres sont constants :

$$\mu_t = \mu, \alpha_t = \alpha \text{ et } \kappa_t = \kappa .$$

2.  $GEV_{(10)}(\mu_t = \mu_1 + \mu_2 Y_t, \alpha, \kappa)$  : le paramètre de position est une fonction linéaire de la covariable.

4.  $GEV_{(11)}(\mu_t = \mu_1 + \mu_2 Y_t, \alpha = \exp(\alpha_1 + \alpha_2 Y_t), \kappa)$  : le paramètre de position et d'échelle sont des fonctions linéaires de la covariable.

5.  $GEV_{(21)}(\mu_t = \mu_1 + \mu_2 Y_t + \mu_3 Y_t^2, \alpha = \exp(\alpha_1 + \alpha_2 Y_t), \kappa)$  : le paramètre de position est une fonction quadratique de la covariable et le paramètre d'échelle est une fonction linéaire de la covariable.



## **SECTION 3 : ARTICLE 2**

### **Application du modèle Log-normal non-stationnaire à des séries de débit**

#### **1 INTRODUCTION**

Dans beaucoup de régions du monde la fréquence des inondations semble avoir augmenté pendant le demi-siècle passé. Plusieurs techniques ont été développées pour prévoir l'écoulement allant des rapports empiriques ou statistiques aux modèles mathématiques détaillés (Sajjad et Slobodan., 2005). L'analyse fréquentielle est l'une des méthodes les plus utilisées pour l'estimation des débits de crue. Cette approche consiste à ajuster un échantillon de données (Débit maximum annuel, Pluie maximale annuelle,...) à une distribution qui représente le mieux la série observée. Une des hypothèses de base de l'analyse fréquentielle est la stationnarité : les caractéristiques statistiques ne dépendent pas du temps. Cependant, plusieurs séries ne vérifient pas cette hypothèse. D'autres modèles sont capables de tenir compte de l'existence d'une tendance, saut ou tout autre type de non-stationnarité en considérant des distributions avec des paramètres qui dépendent de covariables (Coles 2001, El Adlouni *et al.* 2006). En hydrologie, la covariable peut être le temps ou une autre covariable comme un indice climatique.

Aissaoui *et al.* (2007) ont considéré le modèle Log-normal non-stationnaire et l'ont comparé au modèle GEV non-stationnaire. Pour chaque type de modèle non-stationnaire, les modèles GEV et LN qui représentent le mieux ces dépendances ont été comparés par simulation Monte Carlo. Les résultats montrent qu'un bon choix du modèle mène automatiquement à une réduction du biais alors qu'au niveau de l'erreur moyenne quadratique le modèle LN non-stationnaire est toujours meilleur que le modèle GEV non-stationnaire, même dans le cas d'un mauvais choix du modèle.

L'objectif principal du présent travail est d'étudier du maximum annuel des débits journaliers de 17 stations tirées du site de l'UNESCO et qui correspondent à différentes rivières du monde. Les critères de sélections de ces séries sont : (1) la taille de la série qui doit être (supérieure à 70 ans d'observations), (2) les tests d'ajustement montrent que la loi Log-normale représente bien la série étudié. Le modèle Log-normal non-stationnaire est considéré pour l'estimation des débits de crue, pour les 17 stations, par la méthode du maximum de vraisemblance (ML).

## **2 RAPPEL SUR LE MODÈLE LOG-NORMAL NON-STATIONNAIRE**

Les distributions dont les paramètres dépendent de covariables ont été introduites pour l'ajustement des séries non-stationnaires. Différentes formes de non-stationnarité peuvent être observées dans une série de données et différents tests statistiques ont été développés pour les identifier. On s'intéresse surtout à des tendances de type linéaire ou quadratique. On présente dans la section 2.1 le test de Mann Kendall pour tester l'existence d'une tendance, ensuite dans la section 2.2 on présente un rappel sur le modèle Log-normal non-stationnaire avec une dépendance linéaire ou quadratique et la méthode ML pour l'estimation des paramètres.

## 2.1. Test sur la tendance

Afin d'ajuster ce modèle aux séries de débits de pointe des 17 stations étudiées, on test d'abord l'existence d'une tendance pour les 17 séries considérées. Nous avons utilisé le test de Mann-Kendall. Ce test est basé sur la statistique suivante:

$$Z = \sum_{i=1}^{n-1} \sum_{j=i+1}^n \text{sgn}(X_i - X_j) \quad (1)$$

avec

$$\text{sgn}(X_i - X_j) = \begin{cases} 1 & \text{si } X_i > X_j \\ 0 & \text{si } X_i = X_j \\ -1 & \text{si } X_i < X_j \end{cases}$$

Pour une taille de la série étudiée  $n$  qui est assez grande ( $n > 10$ ) la statistique  $Z$  est distribuée suivant une loi Normale avec:

$$E[Z] = \frac{n(n-1)}{4} \quad \text{et} \quad \text{Var}[Z] = \frac{n(n-1)(2n+5)}{72} \quad (2)$$

On a donc la statistique du test

$$S = \frac{Z - E[Z]}{\sqrt{\text{Var}[Z]}} \quad (3)$$

qui est distribuée suivant une loi Normale centrée réduite (Standard).

On rejette l'hypothèse nulle (il n'y a pas de tendance dans la série) si  $S$  calculée est supérieure à  $z_{\alpha/2}$  en valeur absolue.

## 2.2. Le modèle Log-normal non-stationnaire

Le modèle Log-normal non-stationnaire correspond au processus  $X_t$ , tel que pour des valeurs fixes de temps ou/et de covariables il est distribué suivant une loi Log-normal. Ainsi les paramètres sont fonction de temps ou de covariables. Même si (dans le cas des covariables dépendantes du temps) le processus peut être stationnaire, nous employons la même expression de tous ces processus et seront dénotés « distribution LN non-stationnaire  $X \approx LN2(\mu_t, \sigma_t)$  ». Différents genres de dépendance peuvent être considérés en représentant des paramètres comme fonctions des covariables. Cependant, quand deux modèles donnent un bon ajustement, le modèle avec peu de paramètres sera préféré.

Pour étudier l'exactitude des estimations des paramètres, trois modèles sont considérés pour l'ajustement des séries de débit de crue :

$LN_{(0)}(\mu, \sigma)$  qui est un modèle classique où tous les paramètres sont constants  $\mu_t = \mu, \sigma_t = \sigma$ .

$LN_{(1)}(\mu_t = \beta_1 + \beta_2 Y_t, \sigma)$ . Le paramètre  $\mu$  est fonction linéaire d'une covariable  $Y_t$ .

$LN_{(2)}(\mu_t = \beta_1 + \beta_2 Y_t + \beta_3 Y_t^2, \sigma)$ . Le paramètre  $\mu$  est une fonction quadratique d'une covariable temporelle  $Y_t$  et l'autre paramètre est constant.

Notons que, pour la distribution Log-normale, le changement du paramètre  $\mu$  mène à un changement de la moyenne et de la variance de la variable, ce qui correspond aux changements observés des séries hydro-météorologiques.

### 3 CARACTÉRISTIQUES STATISTIQUES DES SÉRIES ÉTUDIÉES

#### 3.1 Résultats du test de Mann-Kendall

La figure 1 présente les séries chronologiques observées à chacune des 17 stations de l'étude. Il s'agit de séries de débit maximum annuel observé dans plusieurs rivières du monde. Ces séries ont été tirées du site de l'UNESCO.

**[Figure 1]**

Afin de vérifier l'existence de tendances dans ces séries, nous avons appliqué le test de Mann-Kendall aux 17 séries. Les résultats du test sont présentés dans le Tableau 1.

**[Tableau 1]**

L'application du test de Mann Kendall montre que parmi les 17 stations, 7 séries semblent avoir une tendance au niveau de signification 5% ( $z_{1-\alpha/2}=1.96$ ). Les 7 stations sont : CSS, CSI, CWS, CRE, CNSPA, HDN et UKTknaturalised. Le modèle Log-normal est ajusté à toutes les séries (les 17 séries retenues), même celles qui ne représentent pas de tendances. L'emploi de plusieurs formes de dépendance (modèles  $LN_{(0)}$ ,  $LN_{(1)}$  et  $LN_{(2)}$ ) permet de représenter le cas stationnaire et deux types de non-stationnarité, linéaire et quadratique. L'emploi d'un test de comparaison des trois modèles permet de choisir le meilleur modèle pour représenter chacune des séries. Le test considéré est celui de la déviance basée sur la fonction de vraisemblance.

Comme il a été mentionné, la taille des séries considérées dans cette étude dépasse 70, ce qui est rare en hydrologie ce qui permet de tirer le maximum d'information sur la forme de la distribution à ajuster ainsi que pour la détection des tendances temporelles.

### **3.2 Caractéristiques statistiques**

Les caractéristiques statistiques des 17 stations sont présentées au tableau 2. Le tableau 2 montre qu'on a une forte variabilité du débit de pointe dans toutes les stations. On remarque aussi que le coefficient d'asymétrie est positif pour toutes les séries, ce qui montre que le modèle LN est un des modèles les plus adéquats pour ce type de séries.

[Tableau 2]

## **4 ESTIMATION DES PARAMÈTRES ET DE LA MÉDIANE**

Les résultats de l'estimation des paramètres par la méthode du maximum de vraisemblance (ML) sont présentés dans l'annexe A.

Les médianes sont calculées conditionnellement à trois valeurs spéciales de la covariable: la valeur minimale, la moyenne et la valeur maximale (Annexe B). La médiane est estimée conditionnellement aux valeurs de la covariable temps pour les trois modèles et pour les 17 stations.

Pour cette application, le choix du modèle est réalisé à l'aide du test de la déviance statistique. Cette déviance représente une méthode simple pour comparer la validité d'un modèle  $M_1$  ou  $M_0$ , sachant que  $M_0 \subset M_1$ , et elle est définie par (Coles, 2001):

$$D = 2\{l_n^*(M_1) - l_n^*(M_0)\} \quad (4)$$

Où  $l_n^*(M)$  est la fonction du log-vraisemblance du modèle  $M$ . La statistique  $D$  est distribuée suivant une loi de chi-deux ( $\chi_\nu^2$ ). Le paramètre  $\nu$  est la différence entre la dimension (le nombre de paramètres) du modèle  $M_1$  et la dimension du modèle  $M_0$ . Des valeurs de  $D$  qui sont plus grandes que les quantiles de la loi de  $\chi_\nu^2$  pour un niveau de confiance particulier, signifie que le modèle  $M_1$  est meilleur que le modèle  $M_0$ .

Les tableaux C.1, C.2 et C.3 de l'annexe C contiennent les résultats du test de la déviance statistique et sa probabilité au non-dépassement pour les 17 stations étudiées.

## 5 DISCUSSION DES RÉSULTATS

Les résultats montrent que pour chaque station il y a un modèle parmi les trois modèles  $LN_{(0)}$ ,  $LN_{(1)}$  et  $LN_{(2)}$  qui s'ajuste aux séries de débit de pointe. À titre d'exemple pour la station Saskatoon il y a une différence significative entre les modèles  $LN_{(1)}$  et  $LN_{(0)}$  puisque  $D=27.3818$  qui est plus grande que le quantile (95%) de la loi  $\chi_1^2$  ( $p = \Pr(\chi_1^2 \leq 27.3818) = 1$ ). D'autre part, pour les modèles  $LN_{(1)}$  et  $LN_{(2)}$  le test statistique montre qu'il y a aussi une différence significative entre ces deux modèles, car  $D=4.1362$  et  $p = \Pr(\chi_1^2 \leq 4.1362) = 0.9580$ . Ainsi, pour le niveau de signification  $1 - \alpha = 95\%$ , le modèle  $LN_{(2)}$  est le plus adéquat pour représenter la dépendance entre le débit de pointe et la covariable temps. La Figure 2 illustre les résultats obtenus à partir des trois modèles (Figure  $LN_{(0)}$  vs  $LN_{(1)}$  et Figure  $LN_{(0)}$  vs  $LN_{(2)}$ ) pour la

station Saskatoon. D'après les résultats du test de la déviance, le modèle  $LN_{(2)}$  était choisi comme étant le plus adéquat pour représenter la série de donnée. On remarque à partir de ces figures que la tendance ainsi que la variabilité de la série sont bien représentées par le modèle  $LN_{(2)}$  et que le cas stationnaire (quand on ignore la tendance) peut conduire à une surestimation des quantiles.

**[Figure 2]**

Les intervalles de confiance sont calculés en générant 5000 ( $N=5000$ ) échantillons par la méthode de Bootstrap paramétrique. Les résultats montrent que le modèle  $LN_{(2)}$  donne le meilleur estimateur de la médiane. D'autre part, la différence entre le cas stationnaire et le cas non-stationnaire est significative, en particulier pour les petites valeurs de la covariable temps. Un résumé de la liste des modèles retenus pour chaque station est présenté dans le Tableau 3.

**[Tableau 3]**

On remarque que, pour les stations où le test de Mann-Kendall montre que le débit de pointe présente une forte tendance (ex:  $S = 4.8481$  pour la station Saskatoon, et  $S = 4.0442$  pour la station IsleMaligne), le modèle  $LN_{(2)}$  s'ajuste bien à la série de données, tandis que pour celles où le débit de pointe ne présente pas de tendance ou une très faible tendance (ex:  $S = 0.5273$  et  $S = 2.0959$ , pour les stations Hope et UK\_Thames\_kingston\_naturalised respectivement) c'est le modèle  $LN_{(0)}$  qui s'ajuste aux données de débit de pointe.

## 6 CONCLUSIONS

Dans le présent travail nous avons ajusté trois modèles non-stationnaires correspondant au modèle Log-normal, à des séries de données des débits de pointe dans différentes rivières du monde. Les modèles non stationnaires sont très flexibles et permettent d'introduire toute sorte de covariables y compris le temps pour décrire les tendances au niveau de la moyenne ou de la variabilité des séries hydro-météorologiques. Le modèle  $LN_{(0)}$  était choisi pour les séries qui présentent une faible tendance, tandis que le modèle  $LN_{(2)}$  a été accepté dans le cas de fortes tendances. La déviance statistique a permis de choisir le meilleur modèle entre les trois modèles qui s'ajuste à la série de données du débit de pointe de chaque station.

Cette application a permis d'illustrer la flexibilité des modèles non-stationnaires d'une manière générale et particulièrement le modèle LN non-stationnaire. Une seule covariable a été considérée pour la détection des tendances afin de tenir compte de cette information dans l'estimation des quantiles et des événements extrêmes. Cependant, d'autres études sont nécessaires pour l'interprétation de la notion de période de retour dans un cadre non-stationnaire.



## RÉFÉRENCES

- Aissaoui-Fqayeh I, El Adlouni S, Ouarda T. BMJ, et St-Hilaire A. (2007). Développement du modèle Log-normal non-stationnaire et comparaison avec le modèle GEV non-stationnaire. Soumis au Journal des Sciences Hydrologiques.
- Coles, G.S. (2001). An Introduction to Statistical Modeling of Extreme Value, Springer, 208 p.
- Davison, A. C. et Ramesh, N. I. (2000). Local likelihood smoothing of sample extremes. Journal of the Royal Statistical Society Series B, Statistics in Society, Volume 62 part 1, pages 191-208.
- El Adlouni S., Ouarda T. BMJ., Zhang X., Roy R. et Bobée B. (2007). Generalized maximum likelihood estimators of the non-stationary GEV model parameters. Water Resources Research. Sous presse.
- Sajjad A et Slobodan P. S. (2005). An artificial neural network model for generating hydrograph from hydro-meteorological parameters. Journal of Hydrology. Vol. 315. pp. 236–251
- Sankarasubramanian, A. et U. Lall (2003). Flood Quantiles and Changing Climate: Seasonal forecasts and causal relations. Water Resources Research, 39(5):1134.



## LISTE DE TABLEAUX

Tableau 1: Résultats du test de Mann-Kendall pour les 17 stations étudiées.

Tableau 2 : Moyenne, Variance et coefficient d'asymétrie des 17 stations étudiées.

Tableau 3 : Le modèle statistique choisi pour chaque station.

## LISTE DE FIGURES

Figure 1: Débit de pointe dans les 17 stations étudiées

Figure 2: La médiane Conditionnelle et ses intervalles de confiance estimés par les modèles  $LN_{(0)}$  vs  $LN_{(1)}$  et les modèles  $LN_{(0)}$  vs  $LN_{(2)}$ .



**Tableau 1. Résultats du test de Mann-Kendall pour les 17 stations étudiées.**

Station	Z	S	E[Z]	$\sigma$ [Z]
Canada_SouthSaskatchewan_Saskatoon (CSS)	805	4.8481	1314	104.9905
Canada_Saguenay_IsleMaligne (CSI)	757	4.0442	1139	94.4572
Canada_Winnipeg_SlaveFalls (CWS)	1621	2.1362	1,39e+003	109.3065
Canada_Assiniboine_Headingley	1247	0.6382	1314	104.9905
Canada_Fraser_Hope	1294	0.5273	1.3505e+003	107.1413
Canada_Red_Emerson (CRE)	1586	2.9945	1278	102.8543
Canada_NorthSaskatchewan_PrinceAlbert (CNSPA)	1048	2.5336	1314	104.9905
Allemagne_Donau(Danube)_Hofkirchen	1771	0.4215	1.8275e+003	134.0538
Canada_Assiniboine_Brandon	1469	0.9223	1580	120.3467
Hongrie_Duna(Danube)_Nagymaros (HDN)	1898	2.4182	2280	157.9662
Lithuania_Neman_Smalininkai	7323	0.3787	7.1825e+003	371.0318
Pologne_Odra_Gozdowice	1749	0.8909	1.8705e+003	136.3858
Roumanie_Danube_Orsova	6029	1.4349	5.5875e+003	307.6964
Russie_Neva_Novosaratovka	3453	1.8011	3875	234.2986
Tchecoslovaquie_Labe_Decin	4181	1.0560	4.4555e+003	259.9521
UK_Thames_kingston_gauged	3886	1.4335	3570	220.4352
UK_Thames_kingston_naturalised(UKTknaturalised)	4032	2.0959	3570	220.4352

**Tableau 2. Moyenne, Variance et Asymétrie des 17 stations étudiées**

<b>Station</b>	<b>Moyenne</b>	<b>Variance</b>	<b>Asymétrie</b>
Canada_SouthSaskatchewan_Saskatoon	751.8904	1.5455e+005	1,1179
Canada_Saguenay_IsleMaligne	3.2234e+003	1.0505e+006	0,4361
Canada_Winnipeg_SlaveFalls	1.3150e+003	2.7945e+005	0,7441
Canada_Assiniboine_Headingley	170.6945	1.3831e+004	0,9521
Canada_Fraser_Hope	7.1009e+003	1.6444e+006	0.7959
Canada_Red_Emerson	417.2042	1.1897e+005	2.0872
Canada_NorthSaskatchewan_PrinceAlbert	655.8082	6.9253e+004	1.6243
Allemagne_Donau(Danube)_Hofkirchen	1.0329e+003	6.9779e+004	0.6302
Canada_Assiniboine_Brandon	122.2925	1.0198e+004	1.3915
Hongrie_Duna(Danube)_Nagymaros	3.7599e+003	7.3736e+005	0.7832
Lithuania_Neman_Smalininkai	1.4623e+003	2.6345e+005	1.0082
Pologne_Odra_Gozdowice	1.0129e+003	1.1932e+005	0.8036
Roumanie_Danube_Orsova	8.8290e+003	4.0705e+006	0.5753
Russie_Neva_Novosaratovka	3.1436e+003	2.1213e+005	0.2289
Tchecoslovaquie_Labe_Decin	726.9328	7.4600e+004	0.6699
UK_Thames_kingston_gauged	327.8542	1.2341e+004	1.2154
UK_Thames_kingston_naturalised	339.7875	1.2248e+004	1.1700

**Tableau 3. Le modèle statistique choisi pour chaque station**

Station	Modèle choisi
Canada_SouthSaskatchewan_Saskatoon	$LN_{(2)}$
Canada_Saguenay_IsleMaligne	$LN_{(2)}$
Canada_Winnipeg_SlaveFalls	$LN_{(1)}$
Canada_Assiniboine_Headingley	$LN_{(0)}$
Canada_Fraser_Hope	$LN_{(0)}$
Canada_Red_Emerson	$LN_{(1)}$
Canada_Assiniboine_Brandon	$LN_{(0)}$
Canada_NorthSaskatchewan_PrinceAlbert	$LN_{(1)}$
Hongrie_Duna(Danube)_Nagymaros	$LN_{(1)}$
Lithuania_Neman_Smalininkai	$LN_{(0)}$
Pologne_Odra_Gozdowice	$LN_{(0)}$
Roumanie_Danube_Orsova	$LN_{(0)}$
Russie_Neva_Novosaratovka	$LN_{(0)}$
Tchecoslovaquie_Labe_Decin	$LN_{(0)}$
UK_Thames_kingston_gauged	$LN_{(0)}$
UK_Thames_kingston_naturalised	$LN_{(0)}$

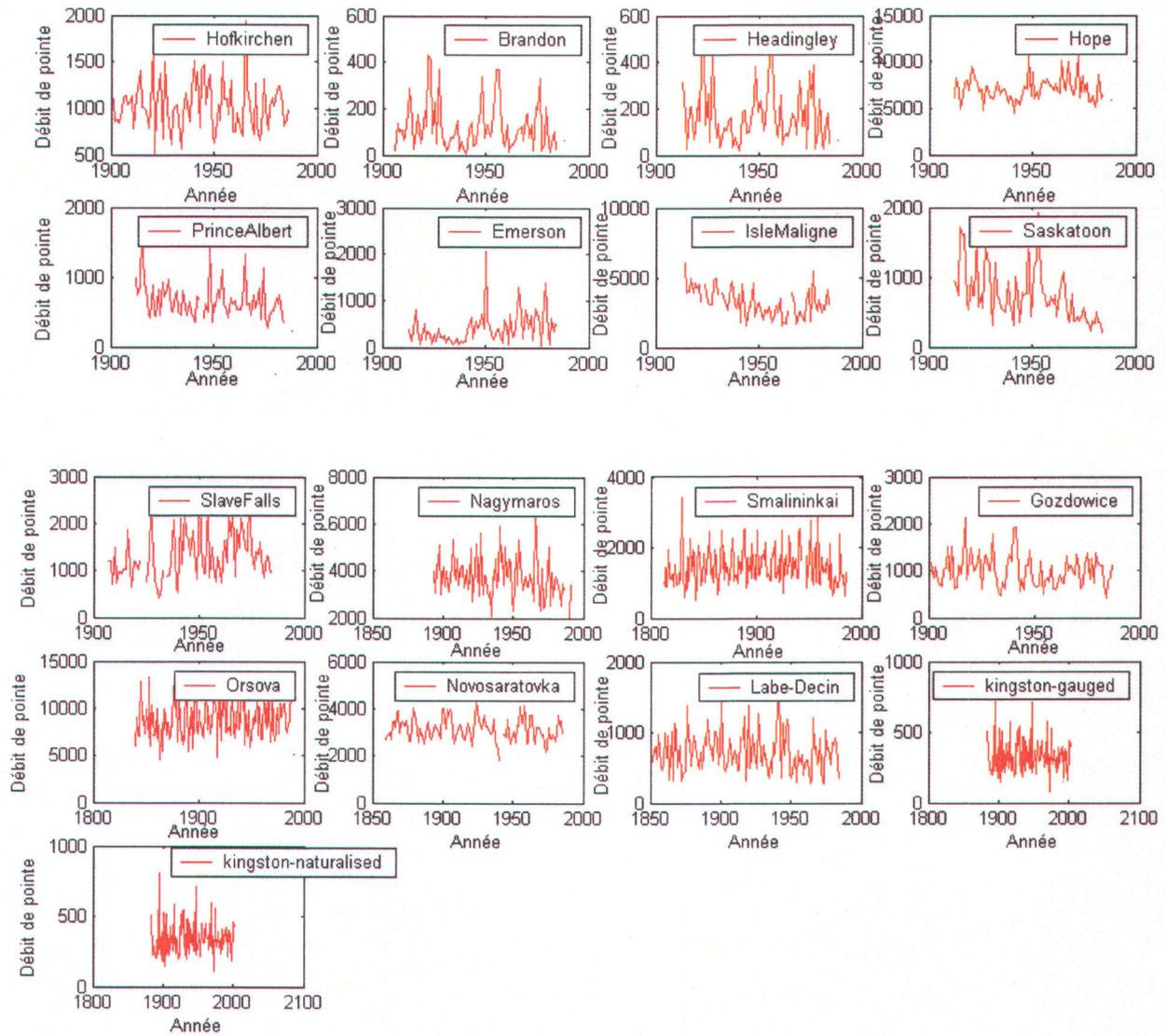
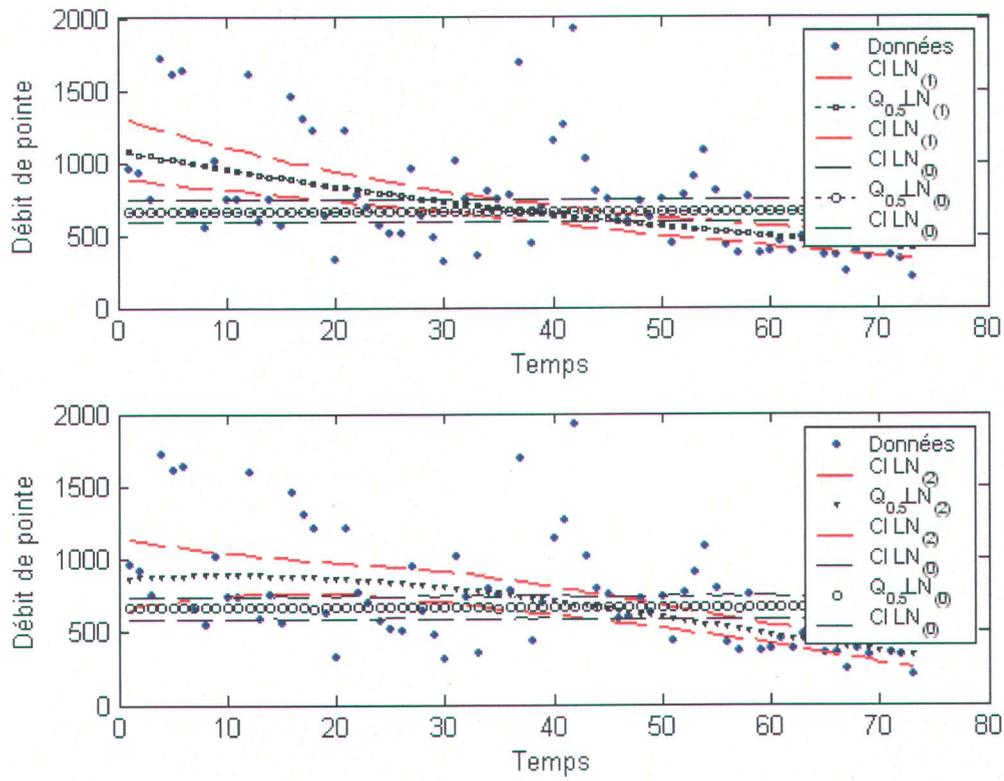


Figure 1. Débit de pointe pour les 17 stations étudiées.



**Figure 2. La médiane Conditionnelle et ses intervalles de confiance estimés par les modèles (en haut)  $LN_{(0)}$  vs  $LN_{(1)}$  et (en bas) les modèles  $LN_{(0)}$  vs  $LN_{(2)}$  pour la station Saskatoon.**



## Annexe A

### A. Les estimateurs des paramètres et les valeurs de la fonction log-vraisemblance

Tableau A.1

		$l_n^*$	$\mu_1$	$\mu_2$	$\mu_3$	$\sigma$
Station Saskatoon	$LN_{(0)}$	-527.3169	6.4974	-	-	0.5003
	$LN_{(1)}$	-513.6260	6.9887	-0.0133	-	0.4147
	$LN_{(2)}$	-511.5579	6.7619	0.0049	-0.0002	0.4032
Station IsleMaligne	$LN_{(0)}$	-565.8191	8.0270	-	-	0.3247
	$LN_{(1)}$	-558.2032	8.2827	-0.0074	-	0.2903
	$LN_{(2)}$	-550.8703	8.5815	-0.0330	0.0004	0.2606
Station Emerson	$LN_{(0)}$	-499.2043	5.7288	-	-	0.8072
	$LN_{(1)}$	-495.3354	5.2763	0.0124	-	0.7650
	$LN_{(2)}$	-495.2986	5.3333	0.0078	0.0001	0.7646
Station SlaveFalls	$LN_{(0)}$	-570.8831	7.1024	-	-	0.4028
	$LN_{(1)}$	-567.6877	6.9003	0.0053	-	0.3860
	$LN_{(2)}$	-567.4490	6.8287	0.0109	-0.0001	0.3848
Station PrinceAlbert	$LN_{(0)}$	-497.6889	6.4181	-	-	0.3609
	$LN_{(1)}$	-494.2725	6.6075	-0.0051	-	0.3444
	$LN_{(2)}$	-494.2679	6.6165	-0.0058	0.00001	0.3444

**Tableau A.2**

		$l_n^*$	$\mu_1$	$\mu_2$	$\mu_3$	$\sigma$
<b>Station Brandon</b>	$LN_{(0)}$	<b>-459.2907</b>	<b>4.4689</b>	-	-	<b>0.8638</b>
	$LN_{(1)}$	<b>-458.6258</b>	<b>4.6634</b>	<b>-0.0048</b>	-	<b>0.8566</b>
	$LN_{(2)}$	<b>-458.4544</b>	<b>4.5333</b>	<b>0.0047</b>	<b>-0.0001</b>	<b>0.8548</b>
<b>Station Headingley</b>	$LN_{(0)}$	<b>-440.1804</b>	<b>4.8815</b>	-	-	<b>0.7631</b>
	$LN_{(1)}$	<b>-439.7869</b>	<b>5.0203</b>	<b>-0.0038</b>	-	<b>0.7590</b>
	$LN_{(2)}$	<b>-439.5400</b>	<b>4.8751</b>	<b>0.0079</b>	<b>-0.0002</b>	<b>0.7564</b>
<b>Station Hope</b>	$LN_{(0)}$	<b>-630.7790</b>	<b>8.8526</b>	-	-	<b>0.1743</b>
	$LN_{(1)}$	<b>-630.7772</b>	<b>8.8504</b>	<b>0.0001</b>	-	<b>0.1743</b>
	$LN_{(2)}$	<b>-630.6799</b>	<b>8.8296</b>	<b>0.0017</b>	<b>-0.00002</b>	<b>0.1740</b>
<b>Station Nagymaros</b>	$LN_{(0)}$	<b>-780.6101</b>	<b>8.2070</b>	-	-	<b>0.2244</b>
	$LN_{(1)}$	<b>-777.7494</b>	<b>8.3015</b>	<b>-0.0019</b>	-	<b>0.2178</b>
	$LN_{(2)}$	<b>-776.6775</b>	<b>8.2268</b>	<b>0.0026</b>	<b>-0.00005</b>	<b>0.2154</b>

**Tableau A.3**

		$l_n^*$	$\mu_1$	$\mu_2$	$\mu_3$	$\sigma$
<b>Station Smalininkai</b>	$LN_{(0)}$	<b>-1286.1</b>	<b>7.2301</b>	-	-	<b>0.3385</b>
	$LN_{(1)}$	<b>-1286</b>	<b>7.2101</b>	<b>0.0002</b>	-	<b>0.3383</b>
	$LN_{(2)}$	<b>-1279.1</b>	<b>6.9938</b>	<b>0.0078</b>	<b>-0.00004</b>	<b>0.3247</b>
<b>Station Gozdowice</b>	$LN_{(0)}$	<b>-625.4988</b>	<b>6.8648</b>	-	-	<b>0.3349</b>
	$LN_{(1)}$	<b>-624.6643</b>	<b>6.9457</b>	<b>-0.0018</b>	-	<b>0.3317</b>
	$LN_{(2)}$	<b>-624.6557</b>	<b>6.9564</b>	<b>-0.0026</b>	<b>0.000008</b>	<b>0.3317</b>
<b>Station Orsova</b>	$LN_{(0)}$	<b>-1348.3</b>	<b>9.0604</b>	-	-	<b>0.2253</b>
	$LN_{(1)}$	<b>-1347.4</b>	<b>9.0177</b>	<b>0.0006</b>	-	<b>0.2240</b>
	$LN_{(2)}$	<b>-1347</b>	<b>8.9781</b>	<b>0.0021</b>	<b>-0.00001</b>	<b>0.2233</b>
<b>Station Novosaratov ka</b>	$LN_{(0)}$	<b>-943.5883</b>	<b>8.0423</b>	-	-	<b>0.1477</b>
	$LN_{(1)}$	<b>-941.7648</b>	<b>8.0861</b>	<b>-0.0007</b>	-	<b>0.1456</b>
	$LN_{(2)}$	<b>-941.6104</b>	<b>8.0695</b>	<b>0.0001</b>	<b>-0.000006</b>	<b>0.1454</b>

Tableau A.4

		$l_n^*$	$\mu_1$	$\mu_2$	$\mu_3$	$\sigma$
Station Decin	$LN_{(0)}$	-935.6520	6.5173	-	-	0.3854
	$LN_{(1)}$	-934.9238	6.5872	-0.0010	-	0.3833
	$LN_{(2)}$	-934.0567	6.4878	0.0033	-0.00003	0.3808
Station kingston_gau kingston_ ged	$LN_{(0)}$	-725.3965	5.7391	-	-	0.3286
	$LN_{(1)}$	-725.0684	5.6967	0.0007	-	0.3277
	$LN_{(2)}$	-724.1357	5.6034	0.0053	-0.00004	0.3252
Station kingston_nat kingston_ uralised	$LN_{(0)}$	-724.7480	5.7790	-	-	0.3141
	$LN_{(1)}$	-723.4208	5.6979	0.0013	-	0.3106
	$LN_{(2)}$	-722.7198	5.6212	0.0051	-0.00003	0.3088
Station Hofkirchen	$LN_{(0)}$	-598.1604	6.9082	-	-	0.2537
	$LN_{(1)}$	-598.1200	6.9219	-0.0003	-	0.2536
	$LN_{(2)}$	-598.1168	6.9167	0.00004	-0.000004	0.2536

## Annexe B

**Tableau B : Les estimateurs des quantiles de probabilité au non-dépassement 50% conditionnellement aux valeurs min, moy, et max de la covariable temps.**

**Tableau B.1**

	Saskatoon			Emerson		
	t=1	t=37	t=73	t=1	t=36.5	t=72
$LN_{(0)}$	663.3976	663.3976	663.3976	307.6134	307.6134	307.6134
$LN_{(1)}$	1070	663.3976	411.3064	198.0827	307.6134	477.7094
$LN_{(2)}$	868.2782	739.6943	333.7660	208.7483	299.3154	503.4311
	IsleMaligne			SlaveFalls		
	t=1	t=34.5	t=68	t=1	t=38	t=75
$LN_{(0)}$	3062.4	3062.4	3062.4	1214.8	1214.8	1214.8
$LN_{(1)}$	3925.5	3062.4	2389.1	997.8402	1214.8	1479
$LN_{(2)}$	5160.7	2654.3	3140.9	934.0594	1257.3	1384.5
	PrinceAlbert			Brandon		
	t=1	t=37	t=73	t=1	t= 40.5	t= 80
$LN_{(0)}$	612.8247	612.8247	612.8247	87.2587	87.2587	87.2587
$LN_{(1)}$	736.8599	612.8247	509.6682	105.4858	87.2587	72.1811
$LN_{(2)}$	743.0055	610.1779	513.9190	93.4966	92.8999	63.9772

**Tableau B.2**

	Headingley			Hope		
	t=1	t=37	t=73	t=1	t=37.5	t=74
$LN_{(0)}$	131.8342	131.8342	131.8342	6992.5	6992.5	6992.5
$LN_{(1)}$	150.8906	131.8342	115.1845	6977.8	6992.5	7007.2
$LN_{(2)}$	132.0037	141.3491	100.7669	6845.3	7062.6	6874.2
	Nagymaros			Smalininkai		
	t=1	t=48.5	t=96	t=1	t= 85.5	t=170
$LN_{(0)}$	3666.7	3666.7	3666.7	1380.4	1380.4	1380.4
$LN_{(1)}$	4022.1	3666.7	3342.7	1353.3	1380.4	1408
$LN_{(2)}$	3749.6	3801.8	3116.2	1098.3	1535.1	1142.7
	Gozdowice			Orsova		
	t=1	t=44	t=87	t=1	t=75.5	t=150
$LN_{(0)}$	957.9368	957.9368	957.9368	8607.6	8607.6	8607.6
$LN_{(1)}$	1036.7	957.9368	885.1224	8252	8607.6	8978.4
$LN_{(2)}$	1047.2	952.9777	894.0418	7944.4	8776	8643.7

**Tableau B.3**

	Novosaratovka			Decin		
	t=1	t= 63	t= 125	t=1	t=67.5	t=134
$LN_{(0)}$	3109.9	3109.9	3109.9	676.7310	676.7310	676.7310
$LN_{(1)}$	3246.7	3109.9	2978.9	724.9930	676.7310	631.6818
$LN_{(2)}$	3195.8	3135.2	2932.2	659.2644	710.4315	574.4129
	kingston gauged			kingston naturalised		
	t=1	t=60.5	t= 120	t=1	t=60.50	t= 120
$LN_{(0)}$	310.7853	310.7853	310.7853	323.4479	323.4479	323.4479
$LN_{(1)}$	298.0976	310.7853	324.0130	298.6418	323.4479	350.3145
$LN_{(2)}$	272.7759	325.2571	296.4900	277.6157	335.7845	325.6504
	Hofkirchen					
	t=1	t= 43.5	t= 86			
$LN_{(0)}$	1000.5	1000.5	1000.5			
$LN_{(1)}$	1013.9	1000.5	987.2715			
$LN_{(2)}$	1009	1003	982.5608			



## Annexe C

Tableau C : Déviance et probabilité au non-dépassement de  $LN_{(0)}/LN_{(1)}$  et  $LN_{(1)}/LN_{(2)}$

**Tableau C.1**

	Saskatoon		Emerson		IsleMaligne	
	Déviance D	Probabilité p	Déviance D	Probabilité p	Déviance D	Probabilité p
$LN_{(0)}/LN_{(1)}$	27.3818	1.0000	7.7378	0.9946	15.2320	0.9999
$LN_{(1)}/LN_{(2)}$	4.1362	0.9580	0.0736	0.2138	14.6656	0.9999
	SlaveFalls		PrinceAlbert		Brandon	
	Déviance D	Probabilité p	Déviance D	Probabilité p	Déviance D	Probabilité p
$LN_{(0)}/LN_{(1)}$	6.3908	0.9885	6.8327	0.9910	1.3297	0.7511
$LN_{(1)}/LN_{(2)}$	0.4773	0.5104	0.0092	0.0765	0.3428	0.4418

**Tableau C.2**

	Headingley		Hope		Nagymaros	
	Déviance D	Probabilité p	Déviance D	Probabilité p	Déviance D	Probabilité p
$LN_{(0)}/LN_{(1)}$	0.7871	0.6250	0.0037	0.0484	5.7214	0.9832
$LN_{(1)}/LN_{(2)}$	0.4937	0.5177	0.1945	0.3408	2.1439	0.8569
	Smalininkai		Gozdowice		Orsova	
	Déviance D	Probabilité p	Déviance D	Probabilité p	Déviance D	Probabilité p
$LN_{(0)}/LN_{(1)}$	0.1967	0.3426	1.6691	0.8036	1.7866	0.8187
$LN_{(1)}/LN_{(2)}$	13.9717	0.9998	0.0170	0.1038	0.9011	0.6575

**Tableau C.3**

	Novosaratovka		Decin		kingston_gauged	
	Déviante D	Probabilité p	Déviante D	Probabilité p	Déviante D	Probabilité p
$LN_{(0)} / LN_{(1)}$	3.6468	0.9438	1.4565	0.7725	0.6561	0.5820
$LN_{(1)} / LN_{(2)}$	0.3089	0.4216	1.7341	0.8121	1.8655	0.8280
	kingston_naturalised		Hofkirchen			
	Déviante D	Probabilité p	Déviante D	Probabilité p		
$LN_{(0)} / LN_{(1)}$	2.6545	0.8967	0.0806	0.2236		
$LN_{(1)} / LN_{(2)}$	1.4020	0.7636	0.0066	0.0646		

## SECTION 4 : CONCLUSION

L'objectif principal de cette étude était de comparer deux modèles non-stationnaires. Le premier correspond à la loi GEV (distribution à trois paramètres) qui est très étudié dans la littérature. Le deuxième modèle correspond à la loi Lognormale (LN, à deux paramètres). Ce dernier modèle a été suggéré pour évaluer l'effet de la forme de la distribution stationnaire sur la flexibilité de ces modèles dans un cadre non-stationnaire, ainsi que l'effet du nombre de paramètres sur les erreurs d'estimation des quantiles conditionnels. La méthode du maximum de vraisemblance (ML) a été considérée pour l'estimation des paramètres des deux modèles.

Les deux modèles, GEV et LN, sont comparés par simulation Monte Carlo, dans le cas de la dépendance linéaire et quadratique à l'égard de la covariable qui est normalement distribuée. La comparaison des deux modèles se base sur le calcul du biais relatif (BR) et la racine de l'erreur quadratique moyenne relative (REQMR) des quantiles. L'étude des propriétés des estimateurs des quantiles des deux modèles montre que les modèles LN non-stationnaires performant bien même dans le cas de séries générées à partir d'un modèle GEV non-stationnaire. En effet, dans la majorité des cas considérés, les REQMR des quantiles estimés par les modèles LN non-stationnaires sont plus petites que celles obtenues par les modèles GEV non-stationnaires. Cependant, le biais est souvent lié au choix de la distribution. On en déduit la flexibilité des modèles LN non-stationnaires pour représenter ce type de dépendance.

L'étude des données de précipitations maximales annuelles, enregistrées à la station Tehachapi de la Californie, conditionnellement à l'indice climatique SOI (article 1), ou des données de débit de pointe conditionnellement au temps (article 2), a été faite en utilisant les trois modèles LN non-stationnaires. En fait, ce type de modèle permet de tenir compte de l'effet d'autres covariables sur

l'évolution de la série observée. Les deux modèles  $LN_{(1)}$  et  $LN_{(2)}$  ont été utilisés pour estimer les quantiles par la méthode ML, et les intervalles de confiance ont été construits par la technique de Bootstrap paramétrique. La comparaison entre les trois modèles, avec un test basé sur la déviance statistique des différents modèles, montre que le modèle  $LN_{(1)}$  est le plus convenable pour représenter la variance des données (article1). En effet, la différence entre les modèles  $LN_{(1)}$  et  $LN_{(2)}$  est non significative, c'est pour cette raison qu'on a choisi le modèle qui a le moins de paramètres. Pour l'article 2 la performance relative des différents modèles diffère d'une station à l'autre.

Les cas étudiés dans ce travail, montrent la flexibilité du modèle LN non-stationnaire pour incorporer l'information additionnelle. En effet, pour l'exemple des précipitations maximales annuelles (article 1), les petites valeurs de la covariable SOI, le modèle classique sous-estime la médiane ce qui peut avoir des conséquences très dramatiques. L'utilisation des estimations conditionnelles a permis de déduire les valeurs des quantiles correspondants aux événements spéciaux qu'on a définis par la covariable.