

**LE TEST GPD ET SON IMPLÉMENTATION
DANS LE LOGICIEL HYFRAN PRO**

**Chaire industrielle Hydro-Québec / CRSNG
en Hydrologie statistique**

Rapport de recherche R-737

mai 2004

**LE TEST GPD ET SON IMPLÉMENTATION
DANS LE LOGICIEL HYFRAN PRO**

Par

**Eva Maria Restle
Salaheddine El Adlouni
Bernard Bobée
et
Taha B. M. J. Ouarda**

**Chaire industrielle Hydro-Québec / CRSNG en Hydrologie statistique
Institut national de la recherche scientifique (INRS)
Centre Eau, Terre et Environnement
2800 rue Einstein CP 7500 Sainte-Foy (Québec) G1V 4C7**

Rapport de recherche R-737

Mai 2004

Référence

RESTLE E.M., EL ADLOUNI S., BOBÉE B. et T.B.M.J. OUARDA (2004). Le test GPD et son implémentation dans le logiciel HYFRAN PRO. INRS-ETE, rapport de recherche R-737, 35 pages.

ÉQUIPE DE RECHERCHE

CHARGÉ DE PROJET

INRS-EAU, TERRE ET ENVIRONNEMENT

Salaheddine El Adlouni

COLLABORATEURS

Eva Maria Restle,
Taha B. M. J. Ouarda

et

Bernard Bobée

Titulaire de la Chaire en Hydrologie statistique

TABLE DES MATIÈRES

1.	INTRODUCTION	1
1.1	CONTEXTE	1
1.2	RÉSUMÉ DE LA MÉTHODE.....	2
2.	CONSTRUCTION, PROPRIÉTÉS ET IMPLÉMENTATION DU TEST GPD.....	5
2.1	CONSTRUCTION DU TEST.....	5
2.2	PROPRIÉTÉS DU TEST.....	9
2.3	IMPLÉMENTATION.....	10
2.3.1	<i>Remarques concernant l'implémentation réalisée</i>	<i>10</i>
2.3.2	<i>Liste des fonctions Matlab créées.....</i>	<i>11</i>
2.3.3	<i>Guide pratique pour l'application de la procédure</i>	<i>14</i>
3.	RÉSULTATS DES SIMULATIONS	17
3.1	LOI GÉNÉRALISÉE DES VALEURS EXTRÊMES.....	18
3.1.1	<i>Indépendance des paramètres</i>	<i>18</i>
3.1.2	<i>Niveau expérimental.....</i>	<i>19</i>
3.2	LOI LOG-NORMALE À 3 PARAMÈTRES.....	22
3.2.1	<i>Indépendance des paramètres</i>	<i>22</i>
3.2.2	<i>Niveau expérimental.....</i>	<i>23</i>
3.3	LOI PEARSON TYPE III	25
3.3.1	<i>Indépendance des paramètres</i>	<i>25</i>
3.3.2	<i>Niveau expérimental.....</i>	<i>26</i>

3.4 LOI HALPHEN TYPE A 29

 3.4.1 *Indépendance des paramètres* 29

 3.4.2 *Niveau expérimental*..... 30

4. **CONCLUSIONS**..... **33**

5. **RÉFÉRENCES** **35**

LISTE DES FIGURES

FIGURE 1.1	ILLUSTRATION DU BUT DE LA PROCÉDURE PRÉSENTÉE. (A) : DISTRIBUTION EMPIRIQUE ET AJUSTÉE, (B) : MODÈLE DE DÉPASSEMENT.	2
FIGURE 3.1	POURCENTAGES D'ACCEPTATION DU VRAI MODÈLE (ESTIMATION DE $1 - \alpha$) POUR τ_3 ENTRE 0.2 ET 0.8 DANS LES CAS DE LA LOI GEV, LES VALEURS DE m_n SONT $m_n = 20$ (A), $m_n = 40$ (B) ET $m_n = 60$ (C).	19
FIGURE 3.2	POURCENTAGES D'ACCEPTATION DU VRAI MODÈLE (ESTIMATION DE $1 - \alpha$) POUR τ_3 ENTRE 0.2 ET 0.8 DANS LE CAS DE LOI LN3, LES VALEURS DE m_n SONT $m_n = 20$ (A), $m_n = 40$ (B) ET $m_n = 60$ (C).	22
FIGURE 3.3	POURCENTAGES D'ACCEPTATION DU VRAI MODÈLE (ESTIMATION DE $1 - \alpha$) POUR τ_3 ENTRE 0.2 ET 0.8 DANS LES CAS DE LA LOI P3, LES VALEURS DE m_n SONT $m_n = 20$ (A), $m_n = 40$ (B) ET $m_n = 60$ (C).	26
FIGURE 3.4	POURCENTAGES D'ACCEPTATION DU VRAI MODÈLE (ESTIMATION DE $1 - \alpha$) POUR τ_3 ENTRE 0.2 ET 0.8 DANS LES CAS DE LA LOI HA, LES VALEURS DE m_n SONT $m_n = 20$ (A), $m_n = 40$ (B) ET $m_n = 60$ (C).	30

LISTE DES TABLEAUX

TABLEAU 3.1	LE TABLEAU DONNE POUR $\alpha = 5\%$ ET CHAQUE COMBINAISON (n, p_n, m_n) LE POURCENTAGE D'ACCEPTATION DE L'HYPOTHÈSE NULLE DANS LE CAS DE LA LOI GEV.....	20
TABLEAU 3.2	VALEURS DE m_n À UTILISER POUR LE TEST GPD DE LA LOI GEV ($\alpha = 5\%$).....	20
TABLEAU 3.3	LE TABLEAU DONNE POUR $\alpha = 1\%$ ET CHAQUE COMBINAISON (n, p_n, m_n) LE POURCENTAGE D'ACCEPTATION DE L'HYPOTHÈSE NULLE DANS LE CAS DE LA LOI GEV.....	21
TABLEAU 3.4	VALEURS DE m_n À UTILISER POUR LE TEST GPD DE LA LOI GEV ($\alpha = 1\%$).....	21
TABLEAU 3.5	LE TABLEAU DONNE POUR $\alpha = 5\%$ ET CHAQUE COMBINAISON (n, p_n, m_n) LE POURCENTAGE D'ACCEPTATION DE L'HYPOTHÈSE NULLE DANS LE CAS DE LA LOI LN3.....	23
TABLEAU 3.6	VALEURS DE m_n À UTILISER POUR LE TEST GPD DE LA LOI LN3 ($\alpha = 5\%$).....	24
TABLEAU 3.7	LE TABLEAU DONNE POUR $\alpha = 1\%$ ET CHAQUE COMBINAISON (n, p_n, m_n) LE POURCENTAGE D'ACCEPTATION DE L'HYPOTHÈSE NULLE DANS LE CAS DE LA LOI LN3.....	24
TABLEAU 3.8	VALEURS DE m_n À UTILISER POUR LE TEST GPD DE LA LOI LN3 ($\alpha = 1\%$).....	25
TABLEAU 3.9	LE TABLEAU DONNE POUR $\alpha = 5\%$ ET CHAQUE COMBINAISON (n, p_n, m_n) LE POURCENTAGE D'ACCEPTATION DE L'HYPOTHÈSE NULLE DANS LE CAS DE LA LOI P3.....	27
TABLEAU 3.10	VALEURS DE m_n À UTILISER POUR LE TEST GPD DE LA LOI P3 ($\alpha = 5\%$).....	27
TABLEAU 3.11	LE TABLEAU DONNE POUR $\alpha = 1\%$ ET CHAQUE COMBINAISON (n, p_n, m_n) LE POURCENTAGE D'ACCEPTATION DE L'HYPOTHÈSE NULLE DANS LE CAS DE LA LOI P3.....	28
TABLEAU 3.12	VALEURS DE m_n À UTILISER POUR LE TEST GPD DE LA LOI P3 ($\alpha = 1\%$).....	28

TABLEAU 3.13	LE TABLEAU DONNE POUR $\alpha = 5\%$ ET CHAQUE COMBINAISON (n, p_n, m_n) LE POURCENTAGE D'ACCEPTATION DE L'HYPOTHÈSE NULLE DANS LE CAS DE LA LOI HA.	31
TABLEAU 3.14	VALEURS DE m_n À UTILISER POUR LE TEST GPD DE LA LOI HA ($\alpha = 5\%$).	31
TABLEAU 3.15	LE TABLEAU DONNE POUR $\alpha = 1\%$ ET CHAQUE COMBINAISON (n, p_n, m_n) LE POURCENTAGE D'ACCEPTATION DE L'HYPOTHÈSE NULLE DANS LE CAS DE LA LOI HA.	32
TABLEAU 3.16	VALEURS DE m_n À UTILISER POUR LE TEST GPD DE LA LOI HA ($\alpha = 1\%$).	32

1. INTRODUCTION

1.1 Contexte

L'analyse fréquentielle a comme but l'estimation des périodes de retour correspondant à une valeur extrême donnée ou de manière équivalente des quantiles extrême correspondant à une période de retour donnée. La méthodologie consiste à ajuster un modèle paramétrique à la distribution des maxima annuels/saisonniers de la variable hydrologique d'intérêt comme par exemple le débit d'une rivière. Un des problèmes rencontrés dans l'application de cette méthodologie est le besoin de choisir un modèle paramétrique parmi une panoplie de modèles possibles et fréquemment utilisés dans le cadre de l'analyse fréquentielle.

Les tests classiques de choix de modèle de distribution, par exemple les tests de Kolmogorov-Smirnov ou de Cramèr-von Mises, testent l'ajustement d'un modèle pour la partie centrale de la distribution des observations. Certaines versions du test de Cramèr-von Mises visent même de diminuer l'influence des observations extrêmes par un schéma de pondération adaptée afin de donner la même importance à chacune des observations. Plusieurs autres tests et critères d'ajustement sont fréquemment utilisés visant chacun l'ensemble de l'échantillon.

Cependant, dans le cadre de l'analyse fréquentielle, nous nous intéressons surtout à l'estimation des événements extrêmes et devons donc étudier d'avantage l'ajustement du modèle choisi pour les queues de la distribution. Malheureusement, il est généralement difficile d'accomplir cet objectif puisque nous disposons très rarement pour ne pas dire jamais de données en nombre suffisant qui correspondent aux événements extrêmes que l'on désire estimer.

Dans le présent rapport, nous présentons un test d'ajustement qui a été spécialement conçu dans le cadre d'estimation d'événements extrêmes. Ce test vise à examiner l'adéquation d'un modèle paramétrique donné aux valeurs extrêmes. L'ouvrage de référence pour la procédure présentée par la suite est la thèse de M. Garrido (Garrido, 2002). La construction du test est basée sur des

résultats asymptotiques de la théorie des valeurs extrêmes (*cf.* par exemple, Embrechts, Klüppelberg et Mikosch, 1997) qui s'applique directement dans le cadre de l'analyse fréquentielle.

La section suivante donne un bref résumé illustratif de la méthode, qui est suivi d'un chapitre fournissant tous les détails techniques et théoriques concernant la construction et l'implémentation de la procédure dans le logiciel `Matlab`. Le chapitre 3 contient les résultats des études de simulation effectuées concernant les propriétés du test proposé et en tire des conclusions. En conclusion, le chapitre 4 résume brièvement les résultats.

1.2 Résumé de la méthode

Dans cette section nous fournissons un bref résumé de l'idée de base et de la construction du test. Le but de la procédure est de tester l'ajustement d'un modèle paramétrique en se concentrant sur la queue de la distribution (*cf.* Figure 1.1 a) et en donnant du poids aux observations extrêmes (*cf.* Figure 1.1 b) :

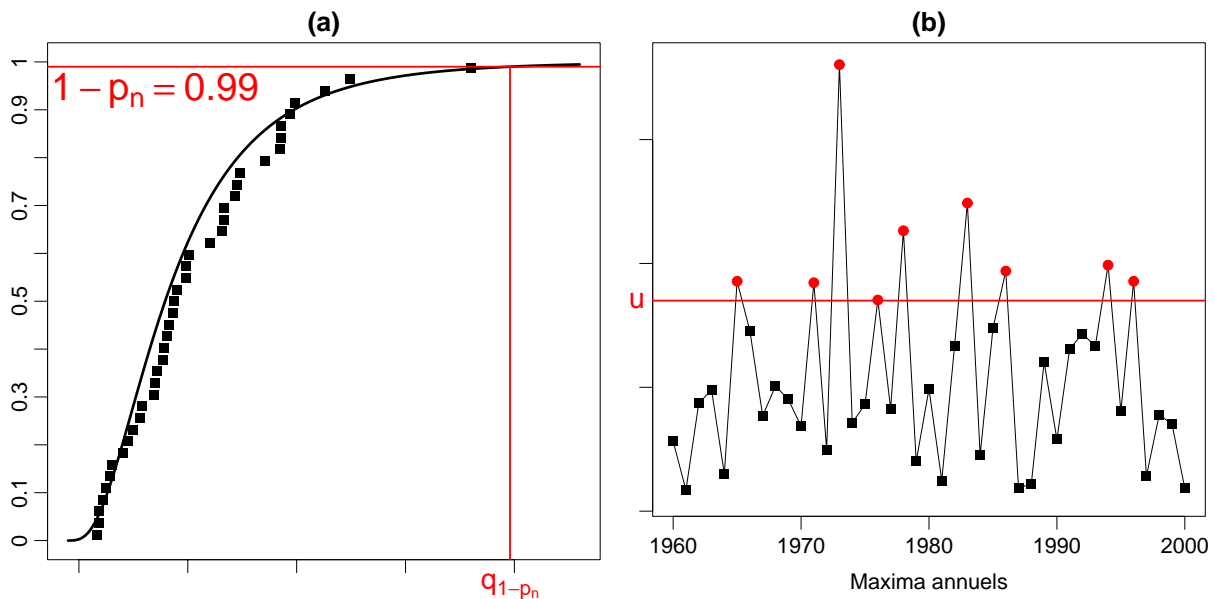


Figure 1.1 Illustration du but de la procédure présentée. (a) : Distribution empirique et ajustée, (b) : Modèle de dépassement.

La base théorique de l'approche est que les excès au-dessus d'un seuil u peuvent être approchés en distribution par la loi de Pareto généralisée. L'idée de base se résume dans les étapes suivantes :

1. Estimer le quantile q_{1-p_n} d'ordre $1-p_n$ en utilisant les excès au-dessus de u et l'approximation Pareto.
2. Construire un intervalle de confiance IC pour q_{1-p_n} à la base de cet estimateur.
3. Accepter le modèle paramétrique si l'estimateur paramétrique $\hat{q}_{\text{param},n}$ appartient à IC .

2. CONSTRUCTION, PROPRIÉTÉS ET IMPLÉMENTATION DU TEST GPD

2.1 Construction du test

Nous présentons dans cette section tous les détails théoriques nécessaires à la construction du test. Pour un échantillon X_1, \dots, X_n provenant d'une loi usuelle F il est possible d'approcher la distribution du maximum de l'échantillon en utilisant la distribution généralisée des valeurs extrêmes :

$$P[M_n = \max(X_1, \dots, X_n) \leq z] \xrightarrow{n \rightarrow \infty} G(z), \text{ où } G(z) = \exp \left\{ - \left[1 + \gamma \left(\frac{z - \mu}{\tilde{\sigma}} \right) \right]^{-1/\gamma} \right\}, \quad (1)$$

défini sur $\{z : [1 + \gamma(z - \mu)/\tilde{\sigma}] > 0\}$ avec $-\infty < \mu < \infty, \tilde{\sigma} > 0$ et $-\infty < \gamma < \infty$.

Cela implique que l'on connaît une approximation de la distribution du maximum d'un échantillon de taille suffisante provenant de n'importe quelle distribution habituellement utilisée en hydrologie. Cependant, en pratique nous n'avons aucune indication précise quant à la taille suffisante de l'échantillon pour rendre cette approximation satisfaisante. Un comportement asymptotique différent est imaginable mais nécessite la construction d'un modèle pour lequel les hypothèses précises (*cf.* par exemple, Embrechts, Klüppelberg et Mikosch 1997; page 158) du théorème ci-dessus ne sont pas vérifiées.

Notons que le cas où $\gamma > 0$ correspond à la distribution de Fréchet, celui où $\gamma < 0$ à la distribution de Weibull et celui où $\gamma \rightarrow 0$ à la distribution de Gumbel. Chacune de ces distributions induit un domaine d'attraction, ce qui veut dire qu'une loi appartenant à un des domaines d'attraction vérifie la relation asymptotique (1) correspondant au signe de γ .

Une conséquence importante de la convergence de la distribution du maximum concerne la distribution des excès d'une variable aléatoire X au-dessus du seuil u (suffisamment grand) qui converge conditionnellement vers la loi de Pareto généralisée :

$$F_u(y) = P[X - u \leq y | X > u] \xrightarrow{n \rightarrow \infty} 1 - \left(1 + \frac{\gamma y}{\sigma}\right)^{-1/\gamma}, \quad (2)$$

défini sur $\{y : y > 0 \text{ et } [1 + \gamma y / \sigma] > 0\}$ avec $\sigma = \tilde{\sigma} + \gamma(u - \mu)$, cf. (Pickands III 1975).

Remarque :

Le cas $\gamma = 0$ est défini dans le sens d'une limite $\gamma \rightarrow 0$, pour le maximum on se retrouve dans le cas d'une distribution asymptotique de Gumbel, et pour les excès dans le cas de la loi exponentielle. Il est possible de faire les développements suivants d'une manière explicite pour le cas de la loi de Gumbel, c'est-à-dire de construire le test pour des échantillons provenant d'une distribution appartenant au domaine d'attraction de Gumbel. Le test résultant s'appelle test ET, où ET est une abréviation de 'exponential tail'. Nous ne poursuivons pas cette piste afin de garantir une plus grande généralité de la procédure développée. De plus, notre démarche est conforme à l'observation que la distribution des excès provenant d'une loi appartenant au domaine d'attraction de Gumbel est souvent mieux approchée par la loi de Pareto généralisée (cf. Garrido, 2002; page 91).

Par la suite nous allons construire l'estimateur du quantile q_{1-p_n} basé sur l'approximation par une loi de Pareto généralisée qui servira pour la définition du test GPD. Le quantile cible d'ordre p_n est défini par $q_{1-p_n} = F^{-1}(1 - p_n)$ pour $1/n < 1 - p_n \leq 1$. Nous utilisons un seuil u_n donné par $u_n = F^{-1}(1 - m_n/n)$, où m_n , le nombre d'excès au-dessus de u_n , est suffisamment petit comparé à n . Alors d'après l'équation (2) :

$$p_n = (1 - F(u_n)) \left(1 - F_{u_n}(q_{1-p_n} - u_n)\right) \approx \frac{m_n}{n} \left(1 + (q_{1-p_n} - u_n) \frac{\gamma}{\sigma}\right)^{-1/\gamma}.$$

En inversant cette égalité, nous trouvons une approximation $q_{\text{GPD},n}$ du quantile cherché :

$$q_{\text{GPD},n} = u_n + \frac{\sigma}{\gamma} \left[\left(\frac{np_n}{m_n} \right)^{-\gamma} - 1 \right].$$

Les termes u_n , γ et σ sont inconnus mais peuvent être estimés : $\hat{u}_n = X_{(n-m_n)}$, les paramètres de la loi de Pareto généralisée, γ et σ , sont estimés par la méthode de Hill généralisé (cf. Dekkers, Einmahl et de Haan, 1989). L'estimateur de q_{1-p_n} est donné par :

$$\hat{q}_{\text{GPD},n} = X_{(n-m_n)} + \frac{\hat{\sigma}_n}{\hat{\gamma}_n} \left[\left(\frac{np_n}{m_n} \right)^{-\hat{\gamma}_n} - 1 \right].$$

Remarque :

L'estimateur de Hill généralisée n'est pas la seule possibilité pour l'estimation de γ et σ . Nous l'avons choisi puisqu'il fournit, dans les situations étudiées, des résultats satisfaisants. Il est cependant possible qu'il faille utiliser une autre méthode pour d'autres situations.

L'erreur d'estimation se décompose de la manière suivante:

$$q_{1-p_n} - \hat{q}_{\text{GPD},n} = \underbrace{q_{1-p_n} - q_{\text{GPD},n}}_{\delta_n, \text{biais d'approximation}} + \underbrace{q_{\text{GPD},n} - \hat{q}_{\text{GPD},n}}_{\text{erreur sur les parametres}}.$$

Soit $F_\theta, \theta \in \Theta$, le modèle paramétrique pour la loi F inconnue que l'on souhaite tester. C'est-à-dire que pour un estimateur $\hat{\theta}_n$ de θ nous souhaitons tester $H_0 : F = F_{\hat{\theta}_n}$ contre $H_1 : F \neq F_{\hat{\theta}_n}$ en prêtant particulièrement attention à l'adéquation dans les queues de $F_{\hat{\theta}_n}$. Le quantile cible estimé à l'aide de ce modèle paramétrique est donné par

$$\hat{q}_{\text{param},n} = F_{\hat{\theta}_n}^{-1}(1-p_n)$$

Le biais δ_n , qui correspond à l'erreur sur le modèle, peut être estimé par $\hat{\delta}_n = \hat{q}_{\text{param},n} - \hat{q}_{\text{GPD},n}$.

Il est possible de construire un intervalle de confiance pour δ_n sur la base de la distribution asymptotique de $\hat{q}_{\text{GPD},n}$, mais les performances du test résultant pour des tailles d'échantillons de

l'ordre de ceux qui nous intéressent en hydrologie ne sont pas satisfaisantes. Nous nous intéressons à deux versions du test GPD basées sur une approche de bootstrap paramétrique : la version bootstrap paramétrique complète et la version bootstrap paramétrique simplifiée du test GPD.

Le **test GPD** dans sa version de **bootstrap paramétrique complète** est effectuée en suivant les étapes ci-dessous :

1. Générer N échantillons indépendants de taille n à partir de $F_{\hat{\theta}_n}$;
2. Calculer N estimations $\hat{q}_{\text{GPD},n}^*$;
3. Calculer N estimations $\hat{\theta}_n^*$ et $\hat{q}_{\text{param},n}^* = F_{\hat{\theta}_n^*}^{-1}(1-p_n)$;
4. Déterminer les quantiles empiriques $\hat{\delta}_{\text{min},n}^*$ et $\hat{\delta}_{\text{max},n}^*$ de probabilité au non dépassement $\alpha/2$ et $1-\alpha/2$ à partir des N valeurs de $\hat{\delta}_n^*$:

$$IC_{\delta, \text{BP}, n} = \left[\hat{\delta}_{\text{min}, n}^*, \hat{\delta}_{\text{max}, n}^* \right];$$

5. Comparer $\hat{\delta}_n$ à $IC_{\delta, \text{BP}, n}$.

Il est possible de construire un **test GPD bootstrap paramétrique simplifié** car les variations d'échantillonnage de $\hat{q}_{\text{param},n}$ sont négligeables par rapport à celles de $\hat{q}_{\text{GPD},n}$. Il consiste à déterminer les quantiles empiriques $\hat{q}_{\text{GPD}, \text{min}, n}^*$ et $\hat{q}_{\text{GPD}, \text{max}, n}^*$ de probabilité au non dépassement $\alpha/2$ et $1-\alpha/2$ à partir des N valeurs de $\hat{q}_{\text{GPD},n}^*$:

$$IC_{\text{GPD}, \text{BP}, n} = \left[\hat{q}_{\text{GPD}, \text{min}, n}^*, \hat{q}_{\text{GPD}, \text{max}, n}^* \right]$$

Étant donné que nous nous intéressons aux échantillons de petite taille nous utilisons la version complète bien qu'elle implique des calculs assez lourds.

Avant l'utilisation et l'implémentation du test, il est nécessaire de déterminer les paramètres suivants :

1. Quantile cible q_{1-p_n} : toujours considérer plusieurs valeur de p_n .
2. Niveau théorique du test α : en général 5%.
3. Nombre d'excès m_n au-dessus du seuil : choisir m_n de sorte que le niveau expérimental du test soit proche de α .

2.2 Propriétés du test

Les propriétés théoriques du test GPD, comme le niveau exact et la puissance du test, ont été étudiées dans la thèse de M. Garrido (Garrido, 2002) et dans un rapport de recherche qu'elle a fournit en supplément pour un premier ensemble de modèles. Les moyens employés pour achever ceci sont des développements asymptotiques des quantités estimées ou une approche par simulation. L'étude du niveau exact du test est nécessaire puisque nous utilisons des intervalles de confiance approximatifs construits à l'aide de la méthode du bootstrap paramétrique.

Pour les modèles auxquels nous nous intéressons nous avons choisi d'employer uniquement l'approche par simulation. Les résultats de ces simulations sont donnés et discutés dans le chapitre 3 du présent rapport.

2.3 Implémentation

Cette section fournit les détails techniques de l'implémentation du test GPD réalisée dans le logiciel `Matlab`.

2.3.1 Remarques concernant l'implémentation réalisée

Les modèles paramétriques que nous avons étudiés sont la loi généralisée des valeurs extrêmes (GEV), la loi log-normale à 3 paramètres (LN3), la loi Pearson type III (P3) et la loi Halphen type A (HA), la paramétrisation utilisée est donnée dans les sections respectives du chapitre 3. Pour chacun de ces modèles nous avons tout d'abord étudié la dépendance du niveau du test en fonction des paramètres du modèle. Le but est d'étudier la dépendance entre le niveau du test et les paramètres du modèle afin d'utiliser la même valeur de m_n indépendamment des valeurs des paramètres estimés. Si pour un futur modèle à étudier, il s'avère que le niveau du test dépend de la valeur des paramètres on ne peut pas conclure que le test GPD n'est pas applicable. Dans une telle situation il faudrait trouver une valeur de m_n à utiliser pour chaque combinaison de paramètres possible, c'est-à-dire au moins sur une grille de combinaisons de paramètres suffisamment fine.

Pour chaque modèle paramétrique étudié nous disposons d'une méthode d'estimation des paramètres associée. Pour les lois GEV, LN3 et P3 nous utilisons la méthode des L-moments (*cf.* Hosking et Wallis, 1997) et pour la loi HA la méthode du maximum de vraisemblance. Cependant, il est possible d'estimer les paramètres par d'autres méthodes d'estimation en effectuant le test GPD.

On rappelle que la méthode employée pour estimer les paramètres γ et σ de la loi GPD est l'estimateur de Hill généralisé. Bien qu'elle fournisse des résultats satisfaisants lorsque elle est utilisée pour le test GPD, nous avons implémenté une deuxième méthode d'estimation basée sur l'approche des L-moments également. Les résultats obtenus indiquent qu'il est tout de même

préférable d'utiliser l'estimateur de Hill pour les modèles paramétriques étudiés dans le présent rapport.

La procédure du bootstrap paramétrique a été généralement répétée 500 fois, ceci a été jugé suffisant pour la précision de l'intervalle de confiance que l'on désire obtenir tout en permettant d'effectuer les simulations nécessaires avec un temps de calcul acceptable. Notons encore une fois que pendant chaque étape du bootstrap les paramètres du modèle sont estimés en utilisant la même méthode que ci-dessus, *i.e.* les L-moments pour tous les modèles sauf pour la loi HA.

2.3.2 Liste des fonctions `Matlab` créées

Cette section fournit une liste de fonctions `Matlab` créées lors de l'implémentation du test GPD. Elles sont triées par leur type d'application. Dans chacune des rubriques nous commençons par les fonctions génériques, c'est-à-dire qui ne dépendent pas du modèle paramétrique. Notons que nous utilisons le logiciel `HYFRAN PRO` pour estimer les paramètres de la loi HA ainsi que pour la génération des vecteurs aléatoires provenant de cette loi.

Fonctions pour l'estimation

- `[sigma, gamma]=GenHillEstimates(data, mn)`
Fournit les estimations de γ et σ des données `data` en utilisant les `mn` observations les plus grandes par la méthode de Hill généralisée.
- `[sigma, gamma]=LmomforGPD(data, numbofexceed)`
Fournit les estimations de γ et σ des données `data` en utilisant les `mn` observations les plus grandes.
- `[qET]=qHatET(data, mn, pn)`
Fournit l'estimation du quantile basée sur l'approximation de queues de distribution avec la loi exponentielle.

- `[qGPD]=qHatGPD(data,mn,pn,method)`
Fournit l'estimation du quantile basée sur l'approximation de queues de distribution avec la loi de Pareto généralisée, la méthode peut être 'Hill' ou 'Lmom'.
- `[l1,l2,l3,l4,t3,t4]=allLmoments(data)`
Fournit les estimations des quatre premiers L-moments ainsi que les L-coefficients d'asymétrie et d'aplatissement correspondant à l'échantillon donné par `data`.
- `br=PWM(data,r)`
Fournit l'estimation du r-ième moment pondéré.
- `[xi,alpha,k]=estimateGEVwithLmoments(l1,t,t3)`
Fournit les estimations des paramètres standard de la loi GEV en fonction des L-moments, `t` correspond à `l2/l1`.
- `[zeta,mu,sigma]=estimateLN3withLmoments(l1,t,t3)`
Fournit les estimations des paramètres standard de la loi LN3 en fonction des L-moments, `t` correspond à `l2/l1`.
- `[alpha,beta,xi]=estimateGenGammawithLmoments(l1,t,t3)`
Fournit les estimations des paramètres standard de la loi P3 en fonction des L-moments, `t` correspond à `l2/l1`.
- `quantile=qGEV(p,xi,alpha,k)`
Fournit l'estimation du quantile d'ordre $1-p$ de la loi GEV en fonction des paramètres standard.
- `quantile = qLN3(p,zeta,mu,sigma)`
Fournit l'estimation du quantile d'ordre $1-p$ de la loi LN3 en fonction des paramètres standard.
- `quantile = qGenGamma(p,xi,beta,alpha)`
Fournit l'estimation du quantile d'ordre $1-p$ de la loi P3 en fonction des paramètres standard.

Fonctions nécessaires pour le bootstrap paramétrique

- `[M, xi, alpha, k]=generateGEVwithLmoments(r, l1, t, t3)`
Génère un vecteur aléatoire de taille r de la loi GEV avec L-moments et L-coefficients correspondant.
- `[M, zeta, mu, sigma]=generateLN3withLmoments(r, l1, t, t3)`
Génère un vecteur aléatoire de taille r de la loi LN3 avec L-moments et L-coefficients correspondant.
- `[M, alphas, betas, xis]=generateGenGammawithLmoments(r, l1, t, t3)`
Génère un vecteur aléatoire de taille r de la loi P3 avec L-moments et L-coefficients correspondant.

Fonctions liées au test GPD

- `[decision, deltan, deltanstarlower, deltanstarupper, deltanstarmean] = GPDtestGEV(data, mn, pn, N, alpha, method)`
Fournit la décision du test GPD pour la loi GEV, 1 si le modèle est accepté 0 sinon, l'estimation du biais δ_n ainsi que les bornes de l'intervalle de confiance et la moyenne de l'échantillon bootstrap de taille N .
- `[decision, deltan, deltanstarlower, deltanstarupper, deltanstarmean] = GPDtestLN3(data, mn, pn, N, alpha, method)`
Fournit la décision du test GPD pour la loi LN3, 1 si le modèle est accepté 0 sinon, l'estimation du biais δ_n ainsi que les bornes de l'intervalle de confiance et la moyenne de l'échantillon bootstrap de taille N .
- `[decision, deltan, deltanstarlower, deltanstarupper, deltanstarmean] = GPDtestGenGamma(data, mn, pn, N, alpha, method)`
Fournit la décision du test GPD pour la loi P3, 1 si le modèle est accepté 0 sinon, l'estimation du biais δ_n ainsi que les bornes de l'intervalle de confiance et la moyenne de l'échantillon bootstrap de taille N .

- `[decision, qGPDstarlower, qGPDstarupper] = GPDtestHA(data, mn, pn, N, alpha, method)`

Fournit la décision du test GPD dans sa version **bootstrap simplifié** pour la loi HA, 1 si le modèle est accepté 0 sinon, ainsi que les bornes de l'intervalle de confiance de l'échantillon bootstrap de taille N . Utilise les fonctions `HYFRAN` et `HYFRAN_HalphenARand` de `HYFRAN PRO` et nécessite que le répertoire de travail de `Matlab` soit celui où se trouve le logiciel `HYFRAN`.

- `[decision, deltan, deltanstarlower, deltanstarupper] = GPDtestHAcomplete(data, mn, pn, N, alpha, method)`

Fournit la décision du test GPD pour la loi HA, 1 si le modèle est accepté 0 sinon, l'estimation du biais δ_n ainsi que les bornes de l'intervalle de confiance et la moyenne de l'échantillon bootstrap de taille N . Utilise les fonctions `HYFRAN` et `HYFRAN_HalphenARand` de `HYFRAN PRO` et nécessite que le répertoire de travail de `Matlab` soit celui où se trouve le logiciel `HYFRAN`.

2.3.3 Guide pratique pour l'application de la procédure

Pour les modèles cités ci-dessus un test GPD pour un échantillon de données s'effectue de la manière suivante :

1. Choisir les valeurs de p_n d'intérêt, choisir le niveau de confiance α .
2. Choisir la valeur de m_n qui correspond à p_n , à α et au modèle que l'on souhaite tester.
3. Effectuer dans `Matlab` la commande correspondant au modèle que l'on souhaite tester, par exemple pour la loi LN3 :

```
GPDtestLN3(data, mn, pn, N, alpha)
```

`Matlab` effectue le test GPD en utilisant la méthode de Hill généralisée pour l'estimation de paramètres de la loi de Pareto généralisée. Si le retour est 1 l'ajustement de la loi LN3 est acceptable en queue de distribution.

Pour d'autres modèles pour lesquels le test GPD n'a pas encore été implémenté il faut effectuer les étapes suivantes :

1. Déterminer la méthode d'estimation ;
2. Fournir un générateur de vecteurs aléatoires provenant du modèle ;
 - ⇒ Implémentation du test GPD en utilisant la méthode de Hill généralisée, la méthode des L-moments ou autre ;
3. Étudier l'indépendance du niveau de test réalisé en fonction des paramètres du modèle ;
4. Effectuer des simulations afin de déterminer les valeurs de m_n à utiliser ;
 - ⇒ Déterminer la méthode d'estimation pour les paramètres de la loi de Pareto généralisée ;
5. Effectuer des simulations concernant la puissance

3. RÉSULTATS DES SIMULATIONS

Ce chapitre donne une synthèse des résultats concernant la performance du test GPD obtenu au moyen de simulations pour chacun des modèles paramétriques étudiés. Par la suite nous donnons d'abord les paramètres des simulations effectuées qui sont les mêmes pour chacun des modèles sauf indication spécifique.

Indépendance

Nous utilisons les coefficients des L-moments pour déterminer le domaine des paramètres d'intérêt dans le cas des modèles de la loi GEV, la loi LN3 et la loi de Pearson type. De plus, nous fixons les paramètres correspondant à la position ($\lambda_1 = 100$) et au L-coefficient de variation ($\tau = 0.2$) et considérons uniquement la variation du paramètre d'asymétrie τ_3 . Celui-ci varie pour des données de débit du Québec essentiellement entre 0.1 et 0.3. Dans l'étude concernant l'indépendance du niveau de test réalisé nous considérons un domaine de variation de 0.2 à 0.8 pour τ_3 . Pour la loi HA les fonctions liant les L-moments aux paramètres de la distribution ne sont pas connus et nous utilisons donc la méthode du maximum de vraisemblance pour l'estimation des paramètres de cette loi. Plus de détails concernant la paramétrisation utilisée se trouvent dans la section 3.4

Les autres paramètres du test GPD sont données par $n = 100$, $N = 500$, $\alpha = 0.05$, $p_n = 1/100$. Le test est effectué 1000 fois pour des valeurs de m_n égales à 20, 40, 60 et le pourcentage d'acceptation du vrai modèle, c'est-à-dire une estimation de $1 - \alpha$, est calculé.

Niveau du test réalisé

Pour les simulations concernant le niveau du test réalisé nous fixons le L-coefficient d'asymétrie $\tau_3 = 0.3$. Les paramètres fixés sont $N = 500$ et $\alpha = 0.05/0.01$, nous effectuons le test 1000 fois pour les valeurs de $n = 30, 50, 100$ et $p_n = 1/100, 1/200, 1/500, 1/1000$, le nombre d'observations utilisés m_n est dans ce cas donné en pourcentage de n et celui-ci varie entre 30% et 70%.

3.1 Loi généralisée des valeurs extrêmes

La paramétrisation habituellement utilisée en hydrologie pour la loi généralisée des valeurs extrêmes diffère légèrement de celle donnée au chapitre 2, sa fonction de distribution s'écrit de la manière suivante :

$$G(z) = \exp \left\{ - \left(1 - \kappa \left(\frac{z - \xi}{\alpha} \right) \right)^{1/\kappa} \right\} \text{ avec } \left[1 - \kappa \frac{z - \xi}{\alpha} \right] > 0,$$

où $-\infty < \xi < \infty$ est le paramètre de position, $\alpha > 0$ le paramètre d'échelle et $-\infty < \kappa < \infty$ le paramètre de forme.

Notons que pour des échantillons provenant de la loi des valeurs extrêmes les excès au-dessus du seuil u sont exactement distribués selon la loi de Pareto généralisée, par conséquent le test GPD aura tendance à accepter la loi GEV comme modèle adéquat ; les résultat pour cette loi sont ainsi à utiliser avec prudence.

3.1.1 Indépendance des paramètres

Les simulations concernant l'indépendance du niveau du test réalisé du paramètre τ_3 sont donnés à la figure ci-dessous. Nous trouvons effectivement une indépendance, du moins pour des valeurs de m_n suffisamment petites et même pour une valeur de $m_n = 60$ et des valeurs de τ_3 souvent rencontrées en hydrologie cette indépendance est vérifiée.

Notons que des valeurs trop élevées de m_n ne sont pas désirable puisque la théorie asymptotique n'est valable que pour des seuils u suffisamment grand correspondant à un nombre faible d'excès.

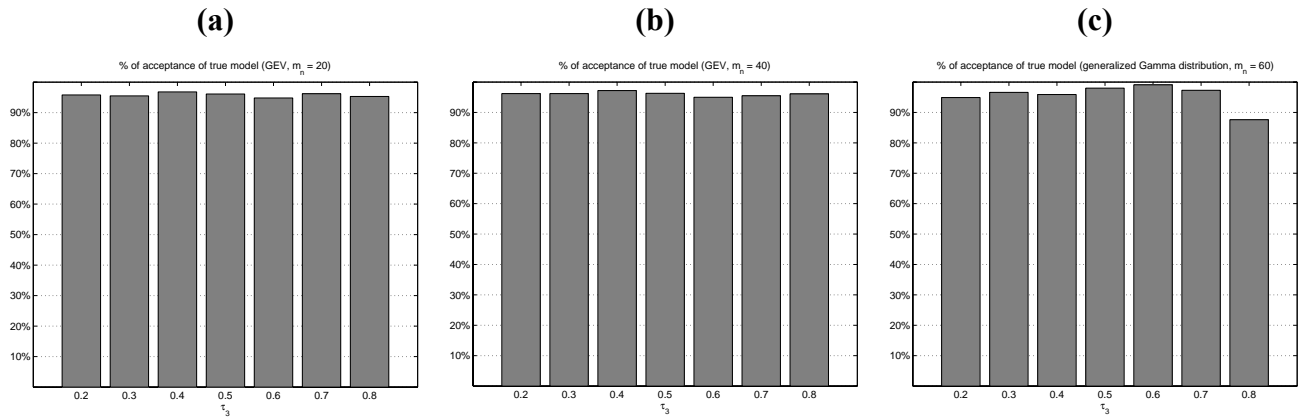


Figure 3.1 Pourcentages d'acceptation du vrai modèle (estimation de $1 - \alpha$) pour τ_3 entre 0.2 et 0.8 dans les cas de la loi GEV, les valeurs de m_n sont $m_n = 20$ (a), $m_n = 40$ (b) et $m_n = 60$ (c).

3.1.2 Niveau expérimental

Les résultats des simulations concernant le niveau du test réalisé sont donnés dans les tableaux 3.1 et 3.3 pour α de 5 et 1 % respectivement. Nous indiquons le pourcentage des tests (sur 1000 tests effectués) qui ne rejettent pas l'hypothèse nulle de la loi GEV alors qu'elle est vraie. Puisque α est la probabilité de rejeter H_0 quand elle est vraie il s'agit donc d'une estimation de $(1 - \alpha)$. Nous sommes maintenant en mesure de choisir une valeur de m_n pour chaque combinaison de n et p_n (cf. Tableaux 3.2 et 3.4).

Tableau 3.1 Le tableau donne pour $\alpha = 5\%$ et chaque combinaison (n, p_n, m_n) le pourcentage d'acceptation de l'hypothèse nulle dans le cas de la loi GEV.

$n = 30$ $\alpha = 5\%$	m_n				
	9	12	15	18	21
$p_n = 1/100$	96.00%	96.45%	97.64%	97.91%	99.27%
$p_n = 1/200$	95.55%	96.18%	97.27%	98.00%	99.09%
$p_n = 1/500$	95.73%	96.27%	97.55%	98.36%	99.18%
$p_n = 1/1000$	95.91%	97.55%	96.82%	98.00%	98.73%
$n = 50$ $\alpha = 5\%$	m_n				
	15	20	25	30	35
$p_n = 1/100$	95.00%	95.27%	97.91%	98.82%	98.27%
$p_n = 1/200$	95.64%	96.45%	96.36%	97.82%	98.82%
$p_n = 1/500$	96.09%	96.73%	97.82%	98.45%	98.73%
$p_n = 1/1000$	96.18%	96.91%	97.55%	97.55%	98.55%
$n = 100$ $\alpha = 5\%$	m_n				
	30	40	50	60	70
$p_n = 1/100$	95.00%	96.18%	97.36%	97.36%	98.27%
$p_n = 1/200$	95.82%	96.82%	97.09%	98.00%	98.64%
$p_n = 1/500$	96.18%	96.18%	96.45%	97.45%	98.64%
$p_n = 1/1000$	95.45%	96.55%	98.09%	98.55%	98.82%

Tableau 3.2 Valeurs de m_n à utiliser pour le test GPD de la loi GEV ($\alpha = 5\%$).

$\alpha = 5\%$	$n = 30$	$n = 50$	$n = 100$
$p_n = 1/100$	9-12	15-20	≈ 30
$p_n = 1/200$	9-12	≈ 15	≈ 30
$p_n = 1/500$	9-12	15-20	30-50
$p_n = 1/1000$	≈ 9	15-20	≈ 30

Tableau 3.3 Le tableau donne pour $\alpha = 1\%$ et chaque combinaison (n, p_n, m_n) le pourcentage d'acceptation de l'hypothèse nulle dans le cas de la loi GEV.

$n = 30$ $\alpha = 1\%$	m_n				
	9	12	15	18	21
$p_n = 1/100$	99.18%	99.09%	99.36%	99.64%	99.82%
$p_n = 1/200$	98.91%	99.45%	99.45%	99.64%	99.64%
$p_n = 1/500$	98.91%	99.27%	99.55%	99.82%	99.91%
$p_n = 1/1000$	98.45%	99.45%	99.55%	99.64%	99.73%
$n = 50$ $\alpha = 1\%$	m_n				
	15	20	25	30	35
$p_n = 1/100$	99.36%	99.45%	99.09%	99.55%	99.82%
$p_n = 1/200$	99.27%	99.45%	99.45%	99.64%	99.82%
$p_n = 1/500$	99.45%	99.00%	99.64%	99.82%	100.00%
$p_n = 1/1000$	98.55%	99.55%	99.55%	99.73%	100.00%
$n = 100$ $\alpha = 1\%$	m_n				
	30	40	50	60	70
$p_n = 1/100$	98.64%	98.91%	99.45%	99.82%	99.45%
$p_n = 1/200$	99.18%	99.09%	99.45%	99.55%	100.00%
$p_n = 1/500$	99.18%	99.55%	99.45%	99.36%	99.73%
$p_n = 1/1000$	99.18%	99.18%	99.45%	99.73%	100.00%

Tableau 3.4 Valeurs de m_n à utiliser pour le test GPD de la loi GEV ($\alpha = 1\%$).

$\alpha = 1\%$	$n = 30$	$n = 50$	$n = 100$
$p_n = 1/100$	9-12	≈ 25	40-50
$p_n = 1/200$	≈ 9	≈ 15	30-40
$p_n = 1/500$	≈ 9	≈ 20	≈ 30
$p_n = 1/1000$	12-15	15-20	30-40

3.2 Loi log-normale à 3 paramètres

La paramétrisation utilisée pour la loi log-normale à 3 paramètres est donnée à l'aide de la fonction de distribution Φ de la distribution normale centrée-réduite :

$$F(x) = \Phi\left(\frac{\log(x-\zeta) - \mu}{\sigma}\right), \text{ avec } x \geq \zeta,$$

où $-\infty < \zeta < \infty$ est le paramètre de position, $\mu > 0$ le paramètre d'échelle et $\sigma > 0$ le paramètre de forme.

3.2.1 Indépendance des paramètres

Les simulations concernant l'indépendance du niveau du test réalisé du paramètre τ_3 sont donnés à la figure ci-dessous. Nous trouvons effectivement une indépendance, du moins pour des valeurs de m_n suffisamment petites et même pour une valeur de $m_n = 60$ et des valeurs de τ_3 souvent rencontrées en hydrologie cette indépendance est vérifiée.

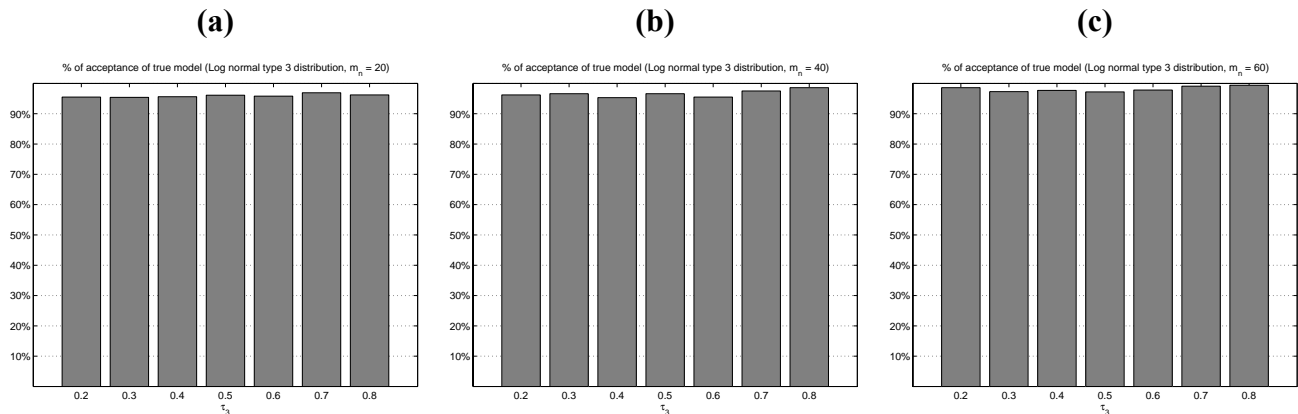


Figure 3.2 Pourcentages d'acceptation du vrai modèle (estimation de $1-\alpha$) pour τ_3 entre 0.2 et 0.8 dans le cas de loi LN3, les valeurs de m_n sont $m_n = 20$ (a), $m_n = 40$ (b) et $m_n = 60$ (c).

3.2.2 Niveau expérimental

Les résultats des simulations concernant le niveau du test réalisé sont donnés dans les tableaux 3.5 et 3.7 pour α de 5 et 1 % respectivement. Nous indiquons le pourcentage des tests (sur 1000 tests effectués) qui ne rejettent pas l'hypothèse nulle de la loi LN3 alors qu'elle est vraie, il s'agit donc d'une estimation de $1 - \alpha$. Nous sommes maintenant en mesure de choisir une valeur de m_n pour chaque combinaison de n et p_n (cf. Tableaux 3.6 et 3.8).

Tableau 3.5 Le tableau donne pour $\alpha = 5\%$ et chaque combinaison (n, p_n, m_n) le pourcentage d'acceptation de l'hypothèse nulle dans le cas de la loi LN3.

$n = 30$ $\alpha = 5\%$	m_n					
	9	12	15	18	21	
	$p_n = 1/100$	95.53%	96.80%	97.16%	98.16%	98.99%
	$p_n = 1/200$	94.99%	96.99%	97.26%	97.71%	98.26%
	$p_n = 1/500$	95.54%	97.27%	96.89%	98.44%	99.17%
	$p_n = 1/1000$	96.45%	96.35%	97.98%	98.89%	99.07%
$n = 50$ $\alpha = 5\%$	m_n					
	15	20	25	30	35	
	$p_n = 1/100$	94.82%	95.91%	97.63%	97.91%	98.18%
	$p_n = 1/200$	94.82%	96.45%	97.00%	97.82%	98.18%
	$p_n = 1/500$	95.18%	95.73%	97.00%	97.73%	99.00%
	$p_n = 1/1000$	96.18%	96.18%	96.82%	97.73%	98.82%
$n = 100$ $\alpha = 5\%$	m_n					
	30	40	50	60	70	
	$p_n = 1/100$	95.09%	95.73%	96.27%	96.73%	98.36%
	$p_n = 1/200$	94.91%	95.09%	97.27%	97.45%	97.73%
	$p_n = 1/500$	95.27%	95.55%	97.09%	97.09%	98.91%
	$p_n = 1/1000$	94.27%	95.73%	96.64%	98.45%	98.91%

Tableau 3.6 Valeurs de m_n à utiliser pour le test GPD de la loi LN3 ($\alpha = 5\%$).

$\alpha = 5\%$	$n = 30$	$n = 50$	$n = 100$
$p_n = 1/100$	≈ 9	15-20	30-40
$p_n = 1/200$	9-12	≈ 15	30-40
$p_n = 1/500$	9-12	15-20	30-40
$p_n = 1/1000$	≈ 9	15-20	30-40

Tableau 3.7 Le tableau donne pour $\alpha = 1\%$ et chaque combinaison (n, p_n, m_n) le pourcentage d'acceptation de l'hypothèse nulle dans le cas de la loi LN3.

$n = 30$ $\alpha = 1\%$	m_n					
	9	12	15	18	21	
	$p_n = 1/100$	99.00%	99.27%	99.36%	99.45%	99.82%
	$p_n = 1/200$	99.09%	99.00%	99.45%	99.54%	99.63%
	$p_n = 1/500$	98.45%	99.82%	99.82%	99.72%	100.00%
	$p_n = 1/1000$	99.27%	98.91%	99.73%	100.00%	99.82%
$n = 50$ $\alpha = 1\%$	m_n					
	15	20	25	30	35	
	$p_n = 1/100$	99.00%	99.18%	99.27%	99.27%	99.82%
	$p_n = 1/200$	99.27%	99.64%	99.18%	99.64%	99.73%
	$p_n = 1/500$	98.45%	99.00%	99.55%	99.73%	99.64%
	$p_n = 1/1000$	98.55%	99.36%	99.45%	99.36%	99.64%
$n = 100$ $\alpha = 1\%$	m_n					
	30	40	50	60	70	
	$p_n = 1/100$	99.00%	98.82%	99.27%	99.18%	99.64%
	$p_n = 1/200$	98.73%	98.64%	99.45%	99.55%	99.36%
	$p_n = 1/500$	98.91%	99.27%	99.45%	99.36%	99.82%
	$p_n = 1/1000$	99.27%	99.36%	99.36%	99.91%	100.00%

Tableau 3.8 Valeurs de m_n à utiliser pour le test GPD de la loi LN3 ($\alpha = 1\%$).

$\alpha = 1\%$	$n = 30$	$n = 50$	$n = 100$
$p_n = 1/100$	≈ 9	15-20	≈ 30
$p_n = 1/200$	9-12	15/25	40-50
$p_n = 1/500$	9-12	≈ 20	30-40
$p_n = 1/1000$	≈ 12	15-20	30-50

3.3 Loi Pearson type III

La paramétrisation utilisée pour la loi Pearson type III est donnée par la fonction de densité :

$$f(x) = \frac{1}{\beta^\alpha \Gamma(\alpha)} (x - \xi)^{\alpha-1} \exp\left(-\frac{x - \xi}{\beta}\right), \text{ avec } x \geq \xi,$$

où $\xi > 0$ est le paramètre de position, $\beta > 0$ le paramètre d'échelle et $\alpha > 0$ le paramètre de forme.

Notons que la loi P3 et la loi LN3 appartiennent au domaine d'attraction maximal de la loi de Gumbel, donc théoriquement il faudrait utiliser le test ET. Puisque la loi P3 a des queues plus légères que la loi LN3 nous avons essayé le test ET pour cette loi et avons trouvé que les résultats ne se sont pas améliorés.

3.3.1 Indépendance des paramètres

Les simulations concernant l'indépendance du niveau du test réalisé du paramètre τ_3 sont donnés à la figure ci-après. Nous trouvons une indépendance pour les valeurs de τ_3 dans le domaine d'intérêt principal, pour des asymétries supérieures à 0.5 elle n'est cependant plus assurée.

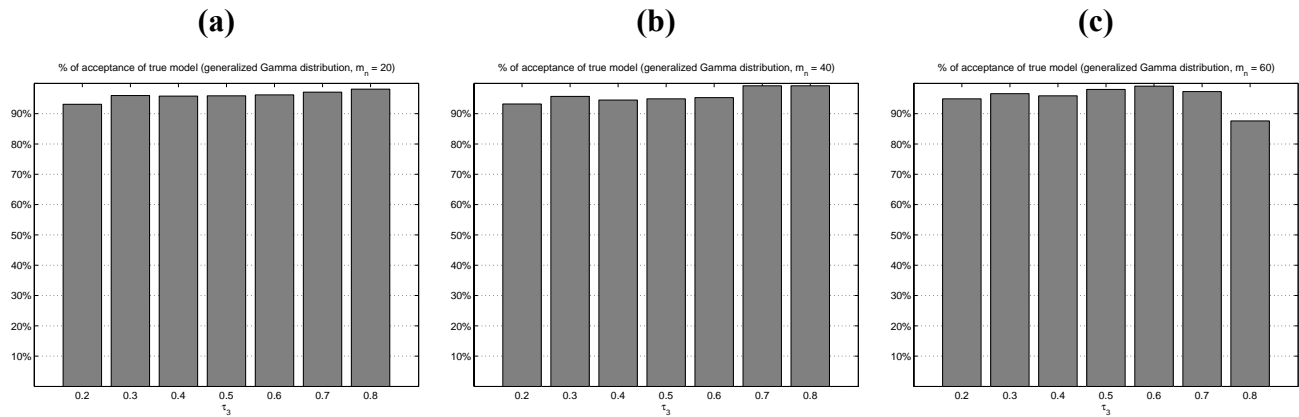


Figure 3.3 Pourcentages d'acceptation du vrai modèle (estimation de $1 - \alpha$) pour τ_3 entre 0.2 et 0.8 dans les cas de la loi P3, les valeurs de m_n sont $m_n = 20$ (a), $m_n = 40$ (b) et $m_n = 60$ (c).

3.3.2 Niveau expérimental

Les résultats des simulations concernant le niveau du test réalisé sont donnés dans les tableaux 3.9 et 3.11 pour α de 5 et 1 % respectivement. Nous indiquons le pourcentage des tests (sur 1000 tests effectués) qui ne rejettent pas l'hypothèse nulle de la loi P3 alors qu'elle est vraie, il s'agit donc d'une estimation de $1 - \alpha$. Nous sommes maintenant en mesure de choisir une valeur de m_n pour chaque combinaison de n et p_n (cf. Tableaux 3.10 et 3.12). Pour la loi P3 le niveau réalisé du test n'est pas satisfaisant pour des échantillons de taille $n = 30$ lorsque $\alpha = 5\%$ ou 1% ni pour des échantillons de taille $n = 50$ lorsque $\alpha = 1\%$, nous recommandons donc d'effectuer le test GPD pour la loi P3 seulement avec des échantillons de taille supérieure.

Tableau 3.9 Le tableau donne pour $\alpha = 5\%$ et chaque combinaison (n, p_n, m_n) le pourcentage d'acceptation de l'hypothèse nulle dans le cas de la loi P3.

$n = 30$ $\alpha = 5\%$	m_n				
	9	12	15	18	21
$p_n = 1/100$	83.52%	85.98%	82.11%	88.04%	85.13%
$p_n = 1/200$	83.71%	84.71%	85.05%	83.64%	86.09%
$p_n = 1/500$	84.44%	84.50%	84.76%	85.48%	83.90%
$p_n = 1/1000$	82.79%	83.18%	85.91%	83.94%	85.39%
$n = 50$ $\alpha = 5\%$	m_n				
	15	20	25	30	35
$p_n = 1/100$	92.18%	93.73%	95.73%	95.64%	96.18%
$p_n = 1/200$	93.73%	93.55%	94.73%	95.18%	96.27%
$p_n = 1/500$	93.64%	94.82%	94.64%	97.00%	96.36%
$p_n = 1/1000$	93.73%	94.09%	94.73%	96.18%	95.82%
$n = 100$ $\alpha = 5\%$	m_n				
	30	40	50	60	70
$p_n = 1/100$	96.00%	95.64%	95.64%	96.82%	98.36%
$p_n = 1/200$	95.09%	95.55%	95.00%	97.09%	97.91%
$p_n = 1/500$	94.73%	95.82%	96.91%	98.18%	98.00%
$p_n = 1/1000$	96.09%	95.36%	96.73%	97.73%	98.00%

Tableau 3.10 Valeurs de m_n à utiliser pour le test GPD de la loi P3 ($\alpha = 5\%$).

$\alpha = 5\%$	$n = 30$	$n = 50$	$n = 100$
$p_n = 1/100$	-	25-35	40-50
$p_n = 1/200$	-	25-30	30-50
$p_n = 1/500$	-	20-25	30-40
$p_n = 1/1000$	-	25/35	≈ 40

Tableau 3.11 Le tableau donne pour $\alpha = 1\%$ et chaque combinaison (n, p_n, m_n) le pourcentage d'acceptation de l'hypothèse nulle dans le cas de la loi P3.

$n = 30$ $\alpha = 1\%$	m_n				
	9	12	15	18	21
$p_n = 1/100$	67.73%	67.64%	68.86%	68.17%	68.78%
$p_n = 1/200$	67.43%	68.56%	68.16%	67.12%	69.24%
$p_n = 1/500$	71.38%	69.89%	66.12%	66.51%	67.22%
$p_n = 1/1000$	70.79%	70.13%	69.37%	68.22%	67.46%
$n = 50$ $\alpha = 1\%$	m_n				
	15	20	25	30	35
$p_n = 1/100$	91.08%	92.81%	93.08%	92.71%	93.72%
$p_n = 1/200$	92.09%	91.82%	93.27%	91.18%	92.55%
$p_n = 1/500$	92.09%	91.91%	93.91%	93.36%	93.82%
$p_n = 1/1000$	92.91%	92.55%	92.55%	94.00%	92.45%
$n = 100$ $\alpha = 1\%$	m_n				
	30	40	50	60	70
$p_n = 1/100$	99.09%	98.27%	98.91%	99.09%	99.73%
$p_n = 1/200$	99.27%	99.27%	98.91%	99.64%	99.64%
$p_n = 1/500$	98.18%	99.36%	98.82%	99.09%	99.55%
$p_n = 1/1000$	98.18%	98.64%	99.09%	99.91%	99.36%

Tableau 3.12 Valeurs de m_n à utiliser pour le test GPD de la loi P3 ($\alpha = 1\%$).

$\alpha = 1\%$	$n = 30$	$n = 50$	$n = 100$
$p_n = 1/100$	-	-	≈ 50
$p_n = 1/200$	-	-	≈ 50
$p_n = 1/500$	-	-	50-60
$p_n = 1/1000$	-	-	≈ 50

3.4 Loi Halphen type A

La paramétrisation utilisée pour la loi Halphen type A est donnée par la fonction de densité :

$$f(x) = \frac{1}{2m^\nu K_\nu(2\phi)} x^{\nu-1} \exp\left(-\phi\left(\frac{x}{m} + \frac{m}{x}\right)\right), \text{ avec } x > 0,$$

où K_ν est la fonction de Bessel de type 2 modifiée, $m > 0$ est le paramètre de position, $\phi > 0$ le premier paramètre de forme $-\infty < \nu < \infty$ le deuxième paramètre de forme. Pour plus d'information concernant la loi HA nous référons à Perreault, Bobée et Rasmussen (1999a) et Perreault, Bobée et Rasmussen (1999b).

Pour la suite nous fixons le paramètre de position ($m_1 = 100$) et considérons uniquement la variation des paramètres de forme. Ceux-ci varient pour des données de débit du Québec essentiellement entre 0 et 8 pour ϕ et -5 et 10 pour ν . Dans l'étude concernant l'indépendance du niveau de test réalisé nous considérons une grille de douze combinaisons des deux paramètres sur ces deux intervalles.

Notons que nous avons utilisé l'implémentation existante des logiciels HYFRAN et HYFRAN PRO pour l'estimation des paramètres et la génération des vecteurs aléatoires provenant de la loi HA. Ceci nous a obligé de supprimer la valeur 1/500 pour p_n puisqu'elle n'est pas prévu dans le lien établie entre HYFRAN et Matlab. Les calculs concernant la loi HA étant lourds nous étions obligés d'utiliser la version simplifiée du test GPD pour ce modèle.

3.4.1 Indépendance des paramètres

Les simulations concernant l'indépendance du niveau du test réalisé des deux paramètres de forme ϕ et ν sont donnés à la figure ci-après. Bien qu'il y ait une variation du pourcentage d'acceptation du vrai modèle, elle n'est pas assez forte ni assez systématique pour en déduire une dépendance du niveau du test des paramètres de forme.

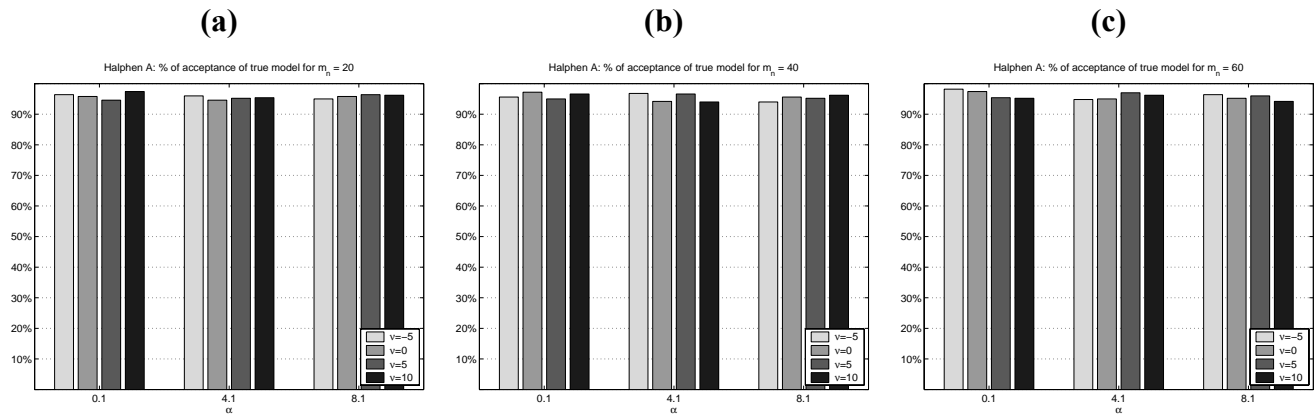


Figure 3.4 Pourcentages d'acceptation du vrai modèle (estimation de $1 - \alpha$) pour τ_3 entre 0.2 et 0.8 dans les cas de la loi HA, les valeurs de m_n sont $m_n = 20$ (a), $m_n = 40$ (b) et $m_n = 60$ (c).

3.4.2 Niveau expérimental

Les résultats des simulations concernant le niveau du test réalisé sont donnés dans les tableaux 3.13 et 3.15 pour α de 5 et 1 % respectivement. Nous indiquons le pourcentage des tests (sur 1000 tests effectués) qui ne rejettent pas l'hypothèse nulle de la loi HA alors qu'elle est vraie, il s'agit donc d'une estimation de $1 - \alpha$. Nous sommes maintenant en mesure de choisir une valeur de m_n pour chaque combinaison de n et p_n (cf. Tableaux 3.14 et 3.16).

Tableau 3.13 Le tableau donne pour $\alpha = 5\%$ et chaque combinaison (n, p_n, m_n) le pourcentage d'acceptation de l'hypothèse nulle dans le cas de la loi HA.

$n = 30$ $\alpha = 5\%$	m_n				
	9	12	15	18	21
$p_n = 1/100$	94.73%	95.45%	95.09%	95.64%	94.73%
$p_n = 1/200$	94.36%	95.64%	95.09%	95.09%	94.36%
$p_n = 1/1000$	94.91%	94.55%	93.27%	96.73%	94.91%
$n = 50$ $\alpha = 5\%$	m_n				
	15	20	25	30	35
$p_n = 1/100$	94.00%	95.45%	95.45%	96.36%	94.00%
$p_n = 1/200$	95.27%	95.64%	96.18%	97.27%	95.27%
$p_n = 1/1000$	94.00%	94.91%	95.09%	95.82%	94.00%
$n = 100$ $\alpha = 5\%$	m_n				
	30	40	50	60	70
$p_n = 1/100$	95.64%	95.64%	95.27%	95.27%	95.64%
$p_n = 1/200$	95.82%	94.73%	96.91%	95.64%	95.82%
$p_n = 1/1000$	93.82%	95.82%	95.45%	94.55%	93.82%

Tableau 3.14 Valeurs de m_n à utiliser pour le test GPD de la loi HA ($\alpha = 5\%$).

$\alpha = 5\%$	$n = 30$	$n = 50$	$n = 100$
$p_n = 1/100$	12-15	20-25	50-60
$p_n = 1/200$	15-18	15-20	≈ 30
$p_n = 1/1000$	18-21	20-25	40-50

Tableau 3.15 Le tableau donne pour $\alpha = 1\%$ et chaque combinaison (n, p_n, m_n) le pourcentage d'acceptation de l'hypothèse nulle dans le cas de la loi HA.

$n = 30$ $\alpha = 1\%$	m_n				
	9	12	15	18	21
$p_n = 1/100$	99.09%	98.73%	98.91%	98.91%	99.09%
$p_n = 1/200$	98.55%	99.45%	98.18%	98.73%	98.55%
$p_n = 1/1000$	98.55%	98.55%	98.55%	99.45%	98.55%
$n = 50$ $\alpha = 1\%$	m_n				
	15	20	25	30	35
$p_n = 1/100$	97.64%	98.73%	99.45%	98.36%	97.64%
$p_n = 1/200$	99.45%	99.09%	98.73%	98.36%	99.45%
$p_n = 1/1000$	98.00%	98.36%	99.64%	98.55%	98.00%
$n = 100$ $\alpha = 1\%$	m_n				
	30	40	50	60	70
$p_n = 1/100$	98.18%	98.91%	98.36%	98.55%	98.18%
$p_n = 1/200$	99.64%	98.36%	99.45%	98.36%	99.64%
$p_n = 1/1000$	98.55%	98.36%	98.36%	98.73%	98.55%

Tableau 3.16 Valeurs de m_n à utiliser pour le test GPD de la loi HA ($\alpha = 1\%$).

$\alpha = 1\%$	$n = 30$	$n = 50$	$n = 100$
$p_n = 1/100$	9/15	20-25	≈ 40
$p_n = 1/200$	9-12	≈ 20	≈ 50
$p_n = 1/1000$	9-15	≈ 30	≈ 60

4. CONCLUSIONS

Dans le présent rapport nous avons présenté une méthode pour tester la qualité de l'ajustement d'un modèle paramétrique à un échantillon en queue de distribution. L'implémentation de la méthode pour quatre modèles paramétriques a été décrite, plus précisément pour la loi GEV, la loi LN3, la loi P3 et la loi HA.

Les simulations présentées au chapitre 4 du rapport ont permis de déterminer quelle version du test GPD est la plus adéquate à utiliser pour chacun des modèles. De plus, elles fournissent une valeur recommandée de m_n pour chaque modèle et chaque combinaison de (n, p_n, α) . Notons que pour la loi P3, la performance du test n'est pas satisfaisante pour des tailles d'échantillon petites. La méthode a été implémentée dans le logiciel HYFRAN PRO, les simulations ont été faites à l'aide du logiciel Matlab.

L'étude du test GPD n'est pas encore complète pour ces quatre modèles. Il reste à considérer la puissance réalisée du test, pour laquelle des premiers résultats ont été obtenus. Ces résultats doivent être validés et comparés aux résultats de Garrido (2002).

5. RÉFÉRENCES

- Dekkers A., J. Einmahl and L. de Haan (1989). A moment estimator for the index of an extreme-value distribution. *The Annals of Statistics* 17(4): 1833-55.
- Embrechts P., C. Klüppelberg and T. Mikosch (1997). Modelling extremal events for insurance and finance. Springer, Berlin; 648 p.
- Garrido M. (2002). Modélisation des événements rares et estimation des quantiles extrêmes : méthodes de sélection de modèles pour les queues de distribution. Thèse de doctorat, Université Joseph Fourier - Grenoble I (France); 231 p.
- Hosking J.R.M. and J.R. Wallis (1997). Regional frequency analysis: An approach based on L-Moments. Cambridge University Press, Cambridge (UK); 224 p.
- Perreault L., B. Bobée and P.F. Rasmussen (1999a). Halphen distribution system - I: Mathematical and statistical properties. *Journal of Hydrology*, 4(3): 189-99.
- Perreault L., B. Bobée and P.F. Rasmussen (1999b). Halphen distribution system - II: Parameter and quantile estimation. *Journal of Hydrology*, 4(3): 200-208.
- Pickands J. (1975). Statistical inference using extreme order statistics. *The Annals of Statistics*, 3(1): 119-31.