

**République Algérienne Démocratique et Populaire**  
**Ministère de l'Enseignement Supérieur et de la Recherche Scientifique**

ECOLE NATIONALE POLYTECHNIQUE

DEPARTEMENT D'HYDRAULIQUE



Choix du modèle fréquentiel adéquat pour  
des chroniques de précipitations maximales  
par minimisation des critères pénalisés AIC  
et BIC

**Projet de Fin d'Etudes**

**Présenté par :**

Amel OUCHERIF

**En vue de l'obtention du diplôme  
d'Ingénieur d'Etat en HYDRAULIQUE**

**Promotion 2012**

# Dédicace et remerciements

---

Je dédie ce modeste travail :

A mes parents qui n'ont cessé de m'accompagner et de me soutenir durant toutes les années de scolarité ;

A mes cher(e)s ami(e)s avec lesquels j'ai passé des moments mémorables, particulièrement lors de la préparation du présent mémoire, l'ambiance qui y régnait était très conviviale. Je leur souhaite beaucoup de réussite dans leur vie professionnelle ;

Mes remerciements à M. Bermad qui a accepté d'encadrer ce travail et de le suivre jusqu'à son aboutissement ;

Mes remerciements également à tous les membres du corps professoral de l'Ecole pour lesquels je suis reconnaissante de m'avoir mené à l'obtention du diplôme d'Ingénieur d'Etat de l'Ecole Nationale Polytechnique ;

Je ne terminerai pas sans remercier mes chers amis du Génie Industriel ainsi que mon camarade Yacine Ait Amiri pour m'avoir aidé à mettre en forme le résumé.

## Résumé:

---

Le présent travail a pour objectif de déterminer la meilleure distribution statistique modélisant les précipitations maximales des différentes séries chronologiques étudiées à savoir la station de Bouira, la station 150106 et la station de Seddouk 151006. Ces dernières appartiennent au bassin versant de la Soummam. En ce qui concerne le bassin versant du Chéelif, les stations considérées sont : la station de Tissemsilt 011006 et celle d'Ain Oussera 011205.

Différentes méthodes ont été utilisées pour la détermination du modèle fréquentiel adéquat en l'occurrence : les ajustements visuels, les diagrammes Quantile-Quantile, les tests d'adéquation, l'harmonisation des temps de retour et des quantiles et enfin la minimisation des critères pénalisés d'Akaike et Bayésien (AIC et BIC). Comparée au reste des méthodes employées, cette dernière s'est révélée être le meilleur moyen pour venir à bout de la problématique du choix du « quasi-vrai » modèle.

**Mots clés :** Critères d'information AIC et BIC, modélisation des chroniques, lois statistiques

## Abstract:

---

The main aim of the present work is to determine the best statistical distribution for modeling maximum rainfall of the different considered chronological series namely: Bouira Station, 150106 Station and Seddouk Station 151006. The latter belong to the Soummam watershed. With regard to the Chéelif watershed, the considered stations are: Tissemsilt 011006 and Ain Oussera 011205.

In order to determine the best frequency model, different methods have been used which are: the visual adjustments, Q-Q plot, goodness of fit tests, the harmonization of the return periods and quintiles and the minimization of Akaike and Bayesian information criteria. Compared to the others methods, the information criteria are the best mean to overcome the problem of choosing the "quasi-true" model.

**Key words :** information criteria AIC and BIC, chronological series modeling, statistical distribution

ملخص :

---

يهدف هذا العمل الى تحديد افضل نموذج إحصائي لتوزيع سقوط الأمطار القصوى للسلاسل الزمنية الآتية: بويرة، سدوق 150106، و151006. هذه المحطات تنتمي الى حوض منحدر الصومام. فيما يخص حوض منحدر الشلف، لدينا السلاسل الزمنية التالية: تسمسيلت 011006، و عين وسارة. 011205

استعملنا عدة طرق و المتمثلة في: التعديلات البصرية، الرسوم البيانية quantile-quantile، اختبارات ملائمة، الموازنة، والمعايير المعلوماتية AIC و BIC.

مقارنة مع الطرق الاخرى، هذه الأخيرة تمثل احسن طريقة لتحديد النموذج الأحسن.

الكلمات المفتاحية: المعايير المعلوماتية AIC و BIC، نمذجة السلاسل الزمنية، القوانين الاحصائية

## Plan de travail

---

Introduction générale.....	1
Chapitre I : Théorème fondamental de la statistique et éléments de base de la statistique descriptive .....	3
I.1 Loi faible des grands nombres.....	3
I.2 Eléments de base de la statistique descriptive .....	4
I.2.1 Moments ordinaires et moments centrés d'ordre $r$ .....	4
I.2.2 Fonction de répartition et densité de probabilité.....	7
I.2.3 Quantiles.....	7
I.2.4 Diagramme Quantile-Quantile : Q-Q plot .....	8
Chapitre II : Principales lois de probabilité et analyse fréquentielle.....	9
II.1 Principales distributions statistiques et leur application pratique en hydrologie .....	9
II.1.1 Lois de probabilité à deux paramètres.....	9
II.1.2 Lois de probabilité à trois paramètres.....	12
II.2 Analyse fréquentielle.....	14
II.2.1 Les différentes formules de calcul de la fréquence empirique.....	15
II.2.2 Notion de temps de retour.....	16
II.2.3 Calcul des quantiles pour les différentes lois de probabilité.....	17
Chapitre III : Eléments clés de la statistique inférentielle.....	22
III.1 Théorie de l'estimation et méthodes d'estimation paramétrique.....	22
III.1.1 Méthode du maximum de vraisemblance.....	24
III.1.2 Méthode des moments.....	27
III.2 Tests d'adéquation.....	28
III.2.1 Test du Khi-deux .....	28
III.2.2 Test de Kolmogorov-Smirnov KS.....	29
III.2.3 Test d'Anderson-Darling.....	29
III.3 Tests d'échantillonnage .....	31
III.3.1 Tests d'homogénéité de Wilcoxon et de Pettitt.....	31

III.3.2 Test de stationnarité KPSS .....	32
III.3.3 Test d'indépendance de Wald-Wolfowitz .....	33
III.3.4 Test des horsains de Grubbs et Beck.....	34
Chapitre IV : Méthodes bayésiennes : critères d'information AIC-BIC.....	35
Chapitre V : Choix du modèle adéquat pour des séries de précipitations maximales.....	39
V.1 Etude des temps de retour et des fréquences empiriques .....	39
V.2 Choix du modèle adéquat pour des séries de précipitations maximales.....	44
V.2.1 Etude de la Station d'Ain Oussera.....	46
V.2.2 Etude de la Station de Bouira .....	57
V.3 Choix de la loi de probabilité adéquate par les critères d'information AIC et BIC.....	69
V.4 Effet du changement de la taille de l'échantillon.....	71
Conclusion générale .....	73

### Introduction générale

Parmi toutes les sciences, la statistique de par toutes ses branches occupe de nos jours, une place très importante dans plusieurs disciplines scientifiques et de recherche en trouvant diverses applications dans différents domaines entre autres l'hydrologie, la médecine, la géophysique, le traitement d'image, la météorologie, les sciences sociales et l'économétrie.

Dans le domaine de l'hydraulique, la combinaison entre la statistique et l'hydrologie est souvent utilisée ; ce qui a donné naissance à l'hydrologie statistique. Celle-ci est une discipline consistant à décrire des processus hydrologiques comme les précipitations ou le ruissellement, en utilisant des techniques statistiques basées essentiellement sur l'analyse des données et les probabilités.

L'une des principales difficultés rencontrées dans ce domaine est la détermination de la loi statistique régissant le phénomène étudié. Cette dernière est d'une importance capitale pour l'ingénieur hydraulicien lors du dimensionnement d'un ouvrage d'art tel que les barrages, où la détermination du quantile correspondant à la crue de projet (de conception) est déterminante. L'ouvrage en question ainsi dimensionné, la loi permettra l'estimation du risque ainsi que la fiabilité de ce dernier et donnera donc une idée très appréciable sur sa durée de vie.

De plus, la connaissance du modèle fréquentiel adéquat devient nécessaire pour l'établissement des courbes Intensité-Durée-Fréquence IDF dans le cas où la loi n'est pas présumée connue (Montana par exemple). En effet, la loi de distribution adéquate doit être ajustée à chaque échantillon pour toutes les durées qu'on se fixe afin de pouvoir calculer les quantiles correspondant aux différentes périodes de retour spécifiées.

Le modèle statistique IDF ainsi établi permettra l'appréhension de la réponse hydrologique (le ruissellement) du bassin versant étudié.

Par ailleurs, l'ajustement approprié permet une analyse du fonctionnement et de la performance hydraulique des réseaux urbains en vue d'éviter leurs débordements causant des inondations et ce, en élaborant une analyse fréquentielle en vue de trouver les valeurs du risque hydrologique lié au dépassement de la pluie de projet, en faisant correspondre à chaque événement critique de période de retour donnée, le risque de débordement. Il devient donc possible de relier la fréquence d'apparition des événements aux dommages éventuels qu'ils peuvent engendrer.

Le choix de la distribution adéquate sert aussi, à cartographier les zones inondables, ce qui permettra aux responsables de prendre des décisions rationnelles quant à la planification de l'occupation des sols (plan d'aménagement du territoire).

En outre, la mise en équation des phénomènes de précipitation et de ruissellement à travers le choix du modèle adéquat permet l'estimation des débits pour un site donné non-jaugé (absence de station de mesure hydrométrique).

Par conséquent, l'importance grandissante prise par le concept du choix du modèle fréquentiel approprié est donc justifiée vu la nécessité de chiffrer la fiabilité et le risque évalués lors de la gestion des projets de conception hydraulique.

Partant de là, l'étude du thème de notre projet de fin d'études sera articulée autour de l'examen des chapitres qui suivent :

Dans le premier chapitre, on commencera par présenter le théorème fondamental de la statistique ainsi que les éléments de base de la statistique descriptive.

Ensuite, dans le deuxième chapitre, nous détaillerons les principales distributions statistiques et l'analyse fréquentielle des données.

Dans le troisième chapitre, nous examinerons les éléments clés de la statistique inférentielle suivis des méthodes bayésiennes en quatrième chapitre.

Enfin, nous terminerons sur le chapitre cinq traitant de la partie pratique de notre travail consistant en le choix du modèle adéquat pour les chroniques de précipitations maximales considérées.

## Chapitre I: Théorème fondamental de la statistique et éléments de base de la statistique descriptive :

La Statistique est la science consistant à déterminer les caractéristiques d'un ensemble de données appelées individus constituant la population étudiée et ce, à travers l'analyse d'un échantillon appartenant à cette même population. Pour ce faire, deux principales approches doivent être adoptées en l'occurrence l'approche exploratoire et l'approche confirmatoire.

La première approche consiste en le traitement de l'échantillon des données collectées en se basant sur la statistique descriptive ou analyse des données afin d'en déterminer les propriétés. Cependant, la seconde approche consiste à utiliser des approches probabilistes de la statistique mathématique ou inférentielle principalement l'estimation et le test afin de confirmer ou infirmer les hypothèses émises sur le comportement de l'échantillon en question.

Par ailleurs, on a tendance à penser, intuitivement, que les observations sont concentrées dans les régions de forte densité de probabilité, contrairement aux zones de faible densité de probabilité où il y a moins d'observations. Cela dit, ceci peut être affecté par le caractère aléatoire du tirage des observations.

Pour mesurer la fidélité de l'échantillon à la distribution, la fonction de répartition empirique est utilisée.

Le théorème fondamental de la Statistique vient confirmer la remarque ci-dessus, celui-ci s'énonçant comme suit :

**« La fonction de répartition empirique  $F_n(x)$  converge vers la fonction de répartition réelle  $F(x)$  quand la taille de l'échantillon tend vers l'infini ».**

En termes mathématiques, cela s'exprime par la loi faible des grands nombres comme suit :

### I.1 Loi faible des grands nombres :

Cette loi confirme le fait intuitif que si une épreuve est répétée un grand nombre de fois, il devient très improbable que la moyenne des  $n$  premières épreuves s'écarte sensiblement de l'espérance du résultat d'une épreuve quand  $n$  augmente sans limites.



- Énoncé de la loi :

Soit  $\{X_i\}$  une suite de variables aléatoires indépendantes et identiquement distribuées (iid) de moyenne commune  $\mu$ .

Pour chaque entier  $n$ , on considère la variable aléatoire  $Y_n$  représentant la moyenne des  $n$  premières  $X_i$ .

$$P\left(\left|\frac{X_1+X_2+\dots+X_n}{n} - \mu\right| < \varepsilon\right) \rightarrow \mathbf{0} \text{ quand } n \rightarrow \infty$$

Mathématiquement, c'est donc la définition d'une limite d'une fonction.

En d'autres termes, il suffit d'augmenter suffisamment  $n$  pour que la moyenne des  $n$  premières  $X_i$  soit très proche de  $\mu$ .

Il s'avère alors que la moyenne empirique est un estimateur convergent (correct) de la moyenne de la population  $\mu$ .

Après avoir énoncé le théorème qui sous-tend la théorie statistique, présentons maintenant les éléments de base de la statistique descriptive.

### **I.2 Éléments de base de la statistique descriptive :**

La statistique descriptive consiste à analyser les séries de données étudiées en les exprimant par des paramètres synthétiseurs particuliers. Trois types de paramètres peuvent être distingués à savoir : les paramètres de tendance centrale, les paramètres de dispersion et les paramètres de forme.

Ces différents paramètres caractérisant une distribution s'expriment par les moments.

Deux types de moments peuvent être distingués : Les moments ordinaires et les moments centrés d'ordre 'n'.

#### **I.2.1 Moments ordinaires et moments centrés d'ordre r :**

##### **I.2.1.1 Moments ordinaires :**

Ces moments sont calculés par rapport à l'origine et ont pour formule :  $E(X^n)$

### *1.2.1.2 Moments centrés :*

Ces moments sont centrés dans le sens où ils sont calculés par rapport à la moyenne. Ainsi, la formule de ce type de moments est la suivante :

$$E[(X - E(X))^n]$$

Notons que les moments centrés d'ordre pair, sont des paramètres de dispersion tandis que ceux d'ordre impair sont des indices d'obliquité ou de dissymétrie.

### *1.2.1.3 Moments remarquables :*

#### *1.2.1.3.1 La moyenne arithmétique :*

La moyenne est un paramètre de tendance centrale représentant le moment ordinaire d'ordre 1, c'est donc l'espérance mathématique en termes de distribution, ou la moyenne arithmétique en termes d'échantillon. Sa formule est donc la suivante :

$$\bar{X} = \frac{1}{n} \sum xi ; \text{ où les } xi \text{ représentent les réalisations de la variable aléatoire } X \text{ et } n \text{ représente la taille de l'échantillon.}$$

#### *1.2.1.3.2 La variance :*

La variance est un paramètre de dispersion, constituant le moment centré d'ordre 2. Il a donc pour formule :

$$E[(X - E(X))^2] = \frac{1}{n-1} \sum (xi - \bar{x})^2 ;$$

#### *1.2.1.3.3 L'écart-type :*

Appelé aussi déviation standard, c'est un paramètre de dispersion qui est égal à la racine de la variance, il a alors pour formule :  $\sqrt{\frac{1}{n-1} \sum (xi - \bar{x})^2}$  ;

Ce paramètre indique la dilatation ou le rétrécissement de la courbe de distribution selon l'importance de sa valeur.

#### *1.2.1.3.4 Le coefficient de dissymétrie :*

Appelé aussi la skewness (notation anglaise), ce coefficient est un moment standardisé qui mesure l'asymétrie de la densité de probabilité d'une variable aléatoire. Pour des distributions symétriques, il est donc nul.

Si sa valeur est positive, cela veut dire que la distribution étudiée est asymétrique à droite, elle sera alors plus étirée à droite. On dira de la dissymétrie qu'elle est gauche.

Dans le cas où ce coefficient est négatif, on dira de la distribution en question qu'elle est asymétrique à gauche, elle sera alors plus étirée à gauche. On dira de la dissymétrie qu'elle est droite.

L'asymétrie représente le troisième moment standardisé, se note par  $C_s$  et a pour expression :

$$C_s = \frac{M_3}{\sigma^3}$$
 où  $M_3$  représente le moment centré d'ordre 3 et  $\sigma^3$  représente le cube de l'écart-type.

Cependant, cette formule représente une forme biaisée de l'asymétrie ; de ce fait, un facteur de correction est affecté à ce coefficient.

Plusieurs études ont été menées dans ce sens, il existe donc plusieurs corrections, entre autres:

$$C_s = \frac{n}{(n-1)(n-2)} \sum \frac{(x_i - \bar{x})^3}{\sigma^3}$$
 ; avec donc le terme  $\left(\frac{n}{(n-1)(n-2)}\right)$  comme facteur de pondération et  $n$  représentant la taille de l'échantillon.

### 1.2.1.3.5 Le coefficient d'aplatissement :

Appelé aussi le kurtosis, ce coefficient sert à mesurer l'amplitude de la distribution étudiée dans le sens où il permet de savoir si la distribution est pointée ou au contraire étalée.

Ainsi, si la valeur de ce coefficient est positive, on dira alors de la distribution qu'elle est leptokurtique. Dans le cas où elle est négative, la distribution est dite platykurtique.

Ce coefficient a pour formule :  $C_k = \left(\frac{M_4}{\sigma^4} - 3\right)$  ;  $M_4$  représentant le moment centré d'ordre 4.

En statistique, un coefficient d'aplatissement négatif s'interprète par une population peu homogène, et un coefficient positif, s'interprète par une population homogène.

Notons que, de même que le coefficient d'asymétrie, il existe plusieurs facteurs de pondération pour réduire le biais.

Enonçons la formule de correction la plus utilisée :

$$C_k = \frac{n(n-1)}{(n-1)(n-2)(n-3)} \sum \frac{(x_i - \bar{x})^4}{\sigma^4} - \frac{3(n-1)^2}{(n-2)(n-3)}$$

### I.2.2 Fonction de répartition et densité de probabilité :

**La fonction de répartition ou de distribution** a pour but de décrire complètement la variable aléatoire étudiée en donnant la valeur de la probabilité pour qu'une réalisation de cette variable soit inférieure à une valeur  $x$ . C'est donc la proportion de la population considérée dont la valeur est inférieure à  $x$ . La formulation mathématique de cette définition peut être définie comme suit :

Soit  $X$  une variable aléatoire. On définit  $F$  la fonction de distribution comme suit :

$$\forall x \in \mathbb{R} : F: \mathbb{R} \rightarrow [0, 1]$$

$$F(x) = P(x \leq X)$$

Les deux événements  $(x \leq X)$  et  $(x > X)$  sont mutuellement exclusifs. Par conséquent :

$$P(x \leq X) + P(x > X) = 1$$

Notons que la fonction de répartition est une fonction continue monotone et non décroissante dont ses valeurs sont comprises entre 0 et 1. Cependant, la fonction de répartition n'est pas nécessairement continue, auquel cas, elle sera en escaliers.

**La fonction de densité de probabilité représente la dérivée de la fonction de répartition.**

### I.2.3 Quantiles :

Appelés aussi fractiles, notés  $F^{-1}(1-P)$  ou  $X_p$  et définis comme suit :

$X_p = \text{Sup} \{x \in \mathbb{R} / F(x) \geq P\}$  où  $P$  correspond à la probabilité au non dépassement.

Ils correspondent donc à la fonction inverse de la fonction de distribution.

### **I.2.4 Diagramme Quantile-Quantile : Q-Q plot :**

Ce diagramme est un outil permettant d'apprécier graphiquement l'ajustement d'un modèle théorique à une série de données observées.

Les valeurs des quantiles  $x_i$  théoriques seront portées sur l'axe des abscisses, tandis que celles des quantiles  $x_i$  observés seront portées sur l'axe des ordonnées.

La distribution théorique testée sera considérée comme un bon ajustement à la série observée si le nuage de points  $(x_i \text{ théo}, x_i \text{ obs})$  suggère une droite représentant la première bissectrice.

Le diagramme Q-Q peut être représenté pour n'importe quelle distribution continue ayant une fonction de répartition strictement croissante et donc bijective.

Une fois le diagramme tracé, si le nuage de points s'aligne sur la première bissectrice alors la distribution théorique représentée s'ajuste bien à l'échantillon étudié.

Une fois toutes les notions précédentes détaillées, on enchaîne avec le chapitre traitant des lois de probabilités les plus utilisées en hydrologie à l'heure actuelle.

## Chapitre II : Principales lois de probabilité et analyse fréquentielle

### II.1 Principales distributions statistiques et leur application pratique en hydrologie :

On appelle série temporelle ou chronologique ou encore chronique, une suite  $(Y_t)$  d'observations chiffrées d'un même phénomène, ordonnées dans le temps.

Une série chronologique peut être décomposée en 3 termes à savoir : un terme exprimant la tendance, un terme exprimant les variations saisonnières, et enfin un terme exprimant les variations résiduelles. Ces dernières sont supposées de faible amplitude.

Dans le cas des chroniques hydrologiques, cas de notre étude, on s'intéressera à la composante tendance, celle-ci correspondant à l'évolution à long terme de la série.

Afin de décrire le comportement statistique des processus hydrologiques, plusieurs modèles fréquentiels peuvent être utilisés selon le type de la variable étudiée.

Dans le présent chapitre, nous allons détailler les principaux modèles que l'on peut utiliser et donner leur application pratique dans le domaine de l'hydrologie.

Nous énoncerons par la suite, l'expression mathématique donnant les quantiles correspondant à chacune des distributions statistiques considérées. Deux types de modèles ont été traités en l'occurrence : les modèles à deux paramètres et les modèles à trois paramètres.

#### II.1.1 Lois de probabilité à deux paramètres :

##### II.1.1.1 Loi Normale et théorème central limite :

**La distribution Normale** est un modèle fréquentiel où les valeurs observées pour une population donnée tournent autour de la moyenne de l'échantillon. Par conséquent, les valeurs s'écartant de cette dernière sont très peu fréquentes.

La loi Normale obéit aux conditions de Borel à savoir qu'elle est générée par un grand nombre de facteurs indépendants dont leurs effets s'additionnent sachant qu'aucun de ces facteurs n'est prédominant.

Les effectifs se distribuant d'une manière sensiblement symétrique par rapport à une valeur centrale, la courbe qui traduit la distribution à savoir la densité de probabilité est alors en cloche où son maximum est, par suite, très proche de la moyenne. Les valeurs aberrantes sont alors négligeables.

Le théorème Central limite permet de montrer que la loi normale représente la somme de plusieurs variables aléatoires. C'est pourquoi, cette loi est souvent utilisée pour la modélisation des totaux et des moyennes. Énonçons donc le théorème Central limite.

**Énoncé du théorème Central limite :** Pour une variable aléatoire quelconque dont l'écart-type est non nul, la moyenne suit asymptotiquement une gaussienne.

Soit :  $(X_1, X_2, \dots, X_n)$  des variables aléatoires indépendantes, toutes suivant la même distribution de moyenne  $\mu$  et de variance  $\sigma^2$ .

En posant la variable :  $Y_n = \sum X_i$ , on aura alors :

$$Z_n = \frac{Y_n - E(Y_n)}{\sqrt{\text{var}(Y_n)}} = \frac{\text{moy}(X) - E(X)}{\sqrt{\text{var}(\text{moy } X)}} \sim N(0,1) \text{ quand } n \rightarrow \infty;$$

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2} \left(\frac{x-\mu}{\sigma}\right)^2\right\} \text{ avec : } \begin{cases} \mu: \text{moyenne} \\ \sigma: \text{écart - type} \end{cases}$$

Pour la loi normale centrée réduite  $N(0,1)$  :

$$f(z) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2} (z)^2\right\} \text{ où } z \text{ représente la variable centrée réduite.}$$

Les valeurs approximatives de la fonction de répartition de la loi normale sont données par la table de Laplace-Gauss.

### **II.1.1.2 Loi Log-Normale :**

La loi log-normale est un cas particulier de la loi normale.

On dit que  $X$  suit une loi log-normale, si son logarithme,  $Y = \ln(X)$ , suit une loi normale.

Les conditions de formation de la loi log-normale sont identiques à celles de la loi normale, à la différence que les facteurs agissent de manière multiplicative et non additive.

Elle est en général utilisée pour la modélisation des extrêmes.

Sa densité de probabilité s'exprime par :

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2} \left(\frac{\ln x - \mu}{\sigma}\right)^2\right\} \text{ avec :}$$

$\left\{ \begin{array}{l} \mu: \text{moyenne liée aux logarithmes des valeurs observées} \\ \sigma: \text{écart - type lié aux logarithmes des valeurs observées} \end{array} \right.$

L'échantillon étudié nécessite donc, dans ce cas, un prétraitement en passant au logarithme des valeurs.

### II.1.1.3 Les lois dérivées de la distribution GEV (General Extreme Values):

Il existe trois lois dérivant de la loi de GEV à savoir : la loi Gumbel, la loi Weibull et enfin la loi Fréchet.

Détaillons chacune de ces lois.

#### II.1.1.3.1 La loi Gumbel :

Appelée aussi loi double exponentielle, sa fonction de répartition a pour formule mathématique :

$$F(x) = e^{-e^{-\frac{(x-a)}{b}}}; \text{ avec : } \left\{ \begin{array}{l} a: \text{paramètre de position} \\ b: \text{paramètre d'échelle (d'étalement)} \end{array} \right.$$

Cette loi est asymétrique et sert à l'estimation des valeurs extrêmes des variables hydrologiques. Elle est très appréciée vu sa simplicité.

#### II.1.1.3.2 La loi Weibull :

Certains expérimentateurs disent que cette loi sert à modéliser les débits d'étiage.

Sa fonction de répartition est de la forme :

$$F(x) = \begin{cases} 1 - e^{-\left(\frac{x}{b}\right)^\lambda}; & x \leq b, \lambda \neq 0; \\ 1; & x \geq b \end{cases}$$

(b,  $\lambda$ ) représentant les paramètres d'échelle et de forme respectivement.



### II.1.1.3.3 La loi Fréchet :

Sa fonction de répartition est de la forme :

$$F(x) = \begin{cases} e^{-\left(\frac{x-a}{b}\right)^{-\lambda}} & ; x > 0, \lambda \neq 0 \\ 0 & ; x \leq 0 \end{cases}$$

### II.1.1.3.4 Loi exponentielle à deux paramètres :

Cette loi constitue un cas particulier de la loi GPD (traitée plus loin). Elle est utilisée pour la modélisation de la distribution des excès au-dessus d'un seuil d'une loi de Gumbel et s'utilise donc pour les séries tronquées.

Sa fonction de répartition est de la forme :

$$F(x) = 1 - e^{-\frac{(x-a)}{b}} ; a \text{ et } b \text{ représentant respectivement les paramètres de position et d'échelle.}$$

## II.1.2 Lois de probabilité à trois paramètres :

### II.1.2.1 Loi des extrêmes généralisée : GEV

Cette loi représente la distribution du maximum de n variables aléatoires. Elle est basée sur le choix d'un bloc temps en prenant une valeur par bloc d'une année ou plus. Elle sert à représenter la distribution du maximum de n variables aléatoires.

Le théorème de Fischer-Tippett constitue la démonstration de cette loi.

La formule mathématique exprimant sa fonction de répartition est la suivante :

$$F(x) = \begin{cases} e^{-\left(1 + \lambda \left(\frac{x-a}{b}\right)\right)^{\frac{-1}{\lambda}}} & ; \lambda \neq 0 \\ e^{-e^{-\frac{(x-a)}{b}}} & ; \lambda = 0 \end{cases}$$

Où : (a, b,  $\lambda$ ) : sont les paramètres de position, d'échelle et de forme respectivement.

On remarque alors que les lois de Gumbel, Fréchet et Weibull sont effectivement des cas particuliers de la GEV, ainsi :

$$\left\{ \begin{array}{l} \text{Si } \lambda = 0, \text{ c'est la loi de Gumbel;} \\ \text{Si } \lambda < 0, \text{ c'est la loi de Weibull;} \\ \text{Si } \lambda > 0, \text{ c'est la loi de Fréchet.} \end{array} \right.$$

### II.1.2.2 Loi Pearson 3 :

Cette loi est un type de loi de probabilité de variables aléatoires réelles positives. En hydrologie, elle sert à la modélisation des crues. C'est une loi très flexible d'un point de vue forme.

Sa fonction de densité de probabilité s'exprime par la formule suivante :

$$F(x) = \frac{|\alpha|}{\Gamma(\lambda)} e^{-\alpha(x-m)^1} (\alpha(x-m))^{\lambda-1}; \lambda > 0 \text{ avec:}$$

$$\left\{ \begin{array}{l} m : \text{paramètre de position} \\ \lambda : \text{paramètre de forme} \\ \alpha : \text{paramètre d'échelle} \end{array} \right.$$

$\Gamma(\lambda) = \int_0^{\infty} e^{-u} u^{\lambda-1} du$  est la fonction mathématique gamma.

Notons qu'en hydrologie, seule la forme à dissymétrie positive c'est-à-dire :

$(\lambda > 0)$  est utilisée car les valeurs manipulées sont positives.

### II.1.2.3 Loi Log-Pearson 3 :

Cette loi est aussi utilisée pour la modélisation des extrêmes et est très flexible d'un point de vue forme. Elle est déduite de la loi de Pearson 3 par une transformation logarithmique. En effet :

Si  $y = \log x$  suit une loi Pearson 3, alors  $x$  suit une loi Log-Pearson 3, dont la densité de probabilité est donnée par :

$$F(x) = \frac{k|\alpha|}{x\Gamma(\lambda)} e^{-\alpha(\log x - m)^1} (\alpha(\log x - m))^{\lambda-1}; \lambda > 0$$

$$k = \frac{1}{\ln a}; \log x : \text{logarithme de base } a ;$$

$$\text{Avec : } \begin{cases} m : \text{paramètre de position} \\ \lambda : \text{paramètre de forme} \\ \alpha : \text{paramètre d'échelle} \end{cases}$$

#### II.1.2.4 Distribution généralisée de Pareto : GPD

Cette loi correspond à la distribution des excès au-dessus d'un seuil d'une loi des GEV. Autrement dit, si les valeurs extrêmes suivent une GEV, alors la GPD agit sur un bloc. Par conséquent, on prendra toutes les valeurs supérieures à un seuil qu'on se donne et par suite toutes les valeurs dépassant ce seuil suivent une GPD. Elle s'avère donc très utile pour la modélisation des séries tronquées.

Sa densité de probabilité est de la forme :

$$F(x) = \begin{cases} 1 - \left(1 - \frac{\lambda}{\alpha(x-u)}\right)^{\frac{1}{\lambda}} ; \lambda \neq 0 \\ 1 - e^{-\frac{x-u}{\alpha}} ; \lambda = 0 \end{cases}$$

Avec :  $(\alpha, u, \lambda)$  représentent les paramètres de position, d'échelle et de forme respectivement.

## II.2 Analyse fréquentielle :

L'hydrologie fréquentielle a pour but de prévoir l'occurrence d'un événement hydrologique dans le futur par le calcul des quantiles et ce, en se basant sur l'interprétation de l'historique des événements hydrologiques d'où l'importance accordée à la disponibilité et la fiabilité de l'information. Quatre étapes essentielles peuvent être distinguées pour le calcul des fréquences :

- Choix d'un échantillon de données satisfaisant certains critères statistiques ;
- La meilleure loi de probabilité théorique reflétant l'échantillon ;
- Inférence statistique pour vérifier l'adéquation de la loi choisie ;
- Utilisation de la loi adéquate pour former un ensemble de scénarios possibles dans le futur.

La distribution statistique possède deux principales caractéristiques à savoir une **capacité descriptive** et une **capacité prédictive**.

La capacité descriptive concerne la capacité du modèle à ajuster adéquatement les données. La capacité prédictive concerne l'estimation des quantiles pour les fortes périodes de retour avec robustesse.

### II.2.1 Les différentes formules de calcul de la fréquence empirique :

Chaque distribution requiert sa propre formule de probabilité empirique, mais il n'existe pas de formules analytiques exactes pour le calcul des fréquences empiriques. Les formules se basant sur la médiane sont indépendantes de la distribution parente des échantillons, et sont utilisées d'une manière standard.

Les valeurs sont classées par ordre croissant ou décroissant ce qui nous donne les fréquences au non dépassement et au dépassement respectivement car par ordre décroissant la valeur de rang  $r$  est toujours dépassée par la valeur de rang  $r-1$  ; même raisonnement pour les fréquences au non dépassement.

En ayant les valeurs classées, la formule générale donnant la fréquence empirique est la suivante :  $F(i) = \frac{i-a}{N+1-2a}$

Enonçons les formules de fréquence empirique les plus utilisées :

- Formule de Californie : Elle constitue l'une des premières formules proposées et a pour formule :  $F(x_i) = \frac{ri}{N}$  ;

Avec :

$\left\{ \begin{array}{l} r_i: \text{rang attribué aux fréquences classées par ordre croissant ou décroissant} \\ N : \text{taille de l'échantillon} \end{array} \right.$

- Pour  $a=0.5$ , on retrouve la formule de Hazen :

$$F(r) = \frac{r-0.5}{N}$$

- Pour  $a=0$ , on retrouve la formule de Weibull :

$$F(r) = \frac{r}{N+1}$$

- Pour  $a=0.3$ , on retrouve la formule de Chegodayev :

$$F(r) = \frac{r-0.3}{N+0.4}$$

- Pour  $a=0.4$ , on retrouve la formule de Cunnane:

$$F(r) = \frac{r-0.4}{N+0.2}$$

- Pour  $a=0.44$ , on retrouve la formule de Gringorten:

$$F(r) = \frac{r-0.44}{N+0.12}$$

A ces formules, on ajoute celle de Tukey:

$$F(r) = \frac{3r-1}{3N+1}$$

### II.2.2 Notion de temps de retour :

Le temps de retour correspond à la moyenne de l'intervalle de récurrence du premier dépassement de la variable étudiée et ce, sur un grand nombre d'occurrence.

Rappelons que l'intervalle de récurrence correspond à la période séparant deux occurrences de l'événement considéré.

Le temps de retour noté  $T$  a pour formule :  $T = \frac{1}{1-F(x)}$  où  $F(x)$  correspond à

la probabilité au non dépassement. Par suite, le temps de retour représente l'inverse de la

probabilité au dépassement correspondant à  $T$ . Ce qui est logique, puisque pour une faible probabilité au dépassement, le temps de retour est fort, ce qui correspond à l'occurrence d'un événement rare.

En termes plus clairs, sur une très longue période, l'événement d'amplitude dépassant ( $x_T$ ) se réalisera en moyenne une fois toutes les  $T$  années. La notion de moyenne est à souligner, car cela implique que la date exacte de l'occurrence de l'événement reste inconnue. C'est pour cela que la notion de temps de retour ne constitue point une notion calendaire mais plutôt probabiliste.

Le calcul du temps de retour pour les différentes distributions sera traité dans la section suivante à savoir le calcul des quantiles, ces deux aspects allant de paire.

### II.2.3 Calcul des quantiles pour les différentes lois de probabilité :

Après avoir choisi le modèle fréquentiel adéquat, nous pourrions alors l'exploiter pour le calcul des quantiles.

Nous allons présenter les formules mathématiques des quantiles correspondant aux différentes distributions citées précédemment dans ce chapitre:

- Loi Normale :

Pour avoir les quantiles pour la loi normale, on utilise l'équation de la droite d'Henri i.e. :

$$x_T = \bar{x} + z\sigma ;$$

( $\bar{x}$ ,  $\sigma$ ,  $z$ ) représentant respectivement la moyenne, l'écart-type des valeurs constituant l'échantillon, et la variable centrée réduite obtenue à partir de la table de Laplace-Gauss selon la fréquence au non dépassement correspondant au temps de retour  $T$  choisi.

Pour obtenir le temps de retour correspondant à une valeur donnée, il suffit de faire l'opération inverse, en commençant par la détermination la variable  $z$  grâce à l'expression suivante déduite de l'équation précédente :

$(z = \frac{x - \bar{x}}{\sigma})$  puis, on tire à partir de la table de Laplace-Gauss, la fréquence au non dépassement(FND) correspondant à z et on obtiendra ainsi le temps de retour voulu par :

$$T = \frac{1}{1 - FND}$$

- Loi Log-Normale :

Pour la loi Log-Normale, la formule mathématique donnant les quantiles est la suivante :

$$\ln x_T = \bar{x} + z\sigma \Rightarrow x_T = \exp\{\bar{x} + z\sigma\};$$

$(\bar{x}, \sigma, z)$  représentant respectivement la moyenne et l'écart-type des logarithmes des valeurs constituant l'échantillon, et la variable obtenue à partir de la table de Laplace-Gauss selon la fréquence au non dépassement correspondant au temps de retour T choisi.

Pour obtenir le temps de retour correspondant à une valeur donnée, il suffit de faire l'opération inverse, en commençant par la détermination la variable z grâce à l'expression suivante déduite de l'équation précédente :

$$(z = \frac{\ln x - \bar{x}}{\sigma}) \text{ puis, on tire à partir de la table de Laplace-Gauss, la fréquence}$$

au non dépassement(FND) correspondant à z et on aura ainsi le temps de retour voulu par :

$$T = \frac{1}{1 - FND}$$

- Loi Gumbel :

Les quantiles correspondant à la fréquence au non dépassement P (ou F) pour cette loi peuvent être obtenus en utilisant directement la fonction de répartition :

$$P = \exp \left[ \exp \left( -\frac{xp - m}{a} \right) \right] \Rightarrow xp = a - b(\ln(-\ln P));$$

Avec : a et b représentant respectivement les paramètres de position, et d'échelle(le gradient exponentiel) estimés de la loi Gumbel.

Pour obtenir le temps de retour correspondant à une valeur donnée, il suffit de calculer la fréquence au non dépassement donnée par la fonction de répartition et obtiendra ainsi le temps de retour voulu par l'expression suivante:  $T = \frac{1}{1-FND}$

- Loi Weibull :

La fonction de répartition de cette distribution s'exprimant par :

$$P = F(x) = 1 - \exp\left[-\left(\frac{x}{b}\right)^\lambda\right]$$

On en déduit que les quantiles  $x_p$  de probabilité au non dépassement  $p$  peuvent être calculés par :  $x_p = b[-\ln(1 - P)]^{1/\lambda}$  ; où  $b$  et  $\lambda$  représentent les paramètres d'échelle et de forme respectivement.

- Loi exponentielle à deux paramètres :

La loi de probabilité de la loi exponentielle étant définie par :

$$F(x) = 1 - \exp\left[-\frac{x-a}{b}\right], \text{ on en déduit ce qui suit :}$$

$x_p = 1 - \exp\left(-\frac{x_p-a}{b}\right)$  ; où  $m$  et  $a$  représentent respectivement les paramètres de position et d'échelle de la loi exponentielle.

- Loi Pearson 3 :

Pour obtenir les quantiles pour la loi Pearson 3, on utilise l'équation générale suivante :

$$x_T = \bar{x} + K\sigma ;$$

$(\bar{x}, \sigma, K)$  représentant respectivement la moyenne et l'écart-type des valeurs constituant l'échantillon, et le facteur de fréquence. Ce dernier a été tabulé pour certaines lois. Pour le cas de la distribution Pearson 3, ce facteur est donné par une table et ce, en fonction du coefficient d'asymétrie et le temps de retour.

Pour obtenir le temps de retour correspondant à une valeur donnée, il suffit de faire l'opération inverse, en commençant par la détermination du facteur de fréquence  $K$  grâce à l'expression suivante déduite de l'équation précédente :



$(K = \frac{x-\bar{x}}{\sigma})$  puis, on tire à partir de la table la valeur de la fréquence au non dépassement(FND) correspondant à K et au coefficient d'asymétrie de la chronique étudiée ; on obtiendra ainsi directement le temps de retour recherché.

Notons que si l'une des valeurs K ou Cs ne figurent pas dans la table, il suffit d'effectuer une interpolation linéaire pour déterminer la valeur recherchée.

- Loi Log-Pearson 3 :

Pour la loi Log-Pearson 3, la formule mathématique donnant les quantiles est la suivante :

$$\text{Ln}x_T = \bar{x} + K\sigma \Rightarrow x_T = \exp\{\bar{x} + K\sigma\} ;$$

$(\bar{x}, \sigma, K)$  représentant respectivement la moyenne et l'écart-type des logarithmes des valeurs constituant l'échantillon, et le facteur de fréquence obtenu à partir de la table de Pearson 3 selon le temps de retour T choisi et le coefficient d'asymétrie Cs de la chronique étudiée.

Pour obtenir le temps de retour correspondant à une valeur donnée, il suffit de faire l'opération inverse, en commençant par la détermination le facteur K grâce à l'expression suivante déduite de l'équation précédente :

$(K = \frac{\text{ln}x-\bar{x}}{\sigma})$  puis, on tire à partir de la table Pearson 3, directement le temps de retour voulu correspondant au coefficient d'asymétrie et au facteur de fréquence calculé.

Notons que dans ce cas, on utilisera le coefficient d'asymétrie de la chronique constituée par le logarithme des valeurs observées.

- Loi GEV :

La fonction de répartition de la loi GEV comme vu précédemment, a pour formule :

$$F(x) = \begin{cases} e^{\left(-\left(1+\lambda\left(\frac{x-a}{b}\right)\right)^{\frac{-1}{\lambda}}\right)}; \\ \lambda \neq 0 \end{cases}$$

Ce qui implique que le quantile  $x_T$  peut être déterminé par la formule suivante :

$$x_T = a - \frac{b}{\lambda} \left(1 - \left(-\ln\left(1 - \frac{1}{T}\right)\right)^{-\lambda}\right)$$

A présent, nous allons enchaîner avec le troisième chapitre traitant des éléments clés de la statistique inférentielle.

## Chapitre III : Éléments clés de la statistique inférentielle

La statistique inférentielle a pour objectif principal de tester l'adéquation du modèle choisi à travers les tests d'hypothèses et de définir les paramètres du modèle choisi à travers les méthodes d'estimation paramétrique.

Cependant, la sélection du modèle est un problème fréquemment rencontré en statistique car dans la plupart des cas, en pratique, les connaissances a priori ne permettent pas de déterminer un unique modèle pour réaliser l'inférence. De ce fait, plusieurs outils mathématiques ont été développés entre autres les tests d'adéquations et les diagrammes d'ajustement des quantiles. Ces derniers représentent des critères classiques pour le choix du modèle adéquat.

Nous allons traiter dans ce qui suit, les aspects d'estimation et de test comme indiqué précédemment.

### III.1 Théorie de l'estimation et méthodes d'estimation paramétrique :

D'un point de vue mathématique, il est possible de déterminer les paramètres standards d'une distribution d'une population pour des variables aléatoires quantitatives. Pour ce faire, il suffit d'avoir la fonction densité de probabilité de la distribution pour en déduire les paramètres  $\mu$  et  $\sigma^2$ .

Mais, dans la pratique, on ne connaît pratiquement jamais exactement la valeur des paramètres précités car la fonction densité de probabilité est très mal connue. Il nous faut donc estimer ces paramètres à partir d'un échantillon de taille « n » tiré au hasard.

On définit alors les paramètres à estimer et leurs estimateurs notés respectivement par  $\theta$  et  $\hat{\theta}$ .

On aura donc, pour notre cas : 
$$\begin{cases} \theta = \mu \text{ ou } \sigma^2 \text{ paramètre à estimer} \\ \hat{\theta} = \hat{\mu} \text{ ou } \widehat{\sigma^2} \text{ estimateur} \end{cases}$$

Notons qu'un bon estimateur doit posséder les propriétés suivantes :

- **Exhaustivité** : En effet, l'estimateur doit être fonction de tous les individus composant l'échantillon tiré ;

- **Absence du biais** : L'estimateur doit être sans biais, c'est-à-dire :  $E(\hat{\theta}) = \theta$  ;

Notons qu'il est possible d'avoir un estimateur asymptotiquement sans biais et ce, pour une valeur de « n » tendant vers l'infini.

- **Convergence** : Un estimateur est dit convergent si la variance de l'estimateur tend vers zéro pour « n » tendant vers l'infini, on écrit alors :

$$\lim_{n \rightarrow \infty} V(\hat{\theta}) = 0 ;$$

- **Efficacité** : C'est un critère permettant de choisir entre deux estimateurs en se basant sur la MSE (Mean Squared Error) ou Moyenne des Carrés de l'Erreur définie par :

$$MSE(\hat{\theta}) = E(\hat{\theta} - \theta)^2 = V(\hat{\theta}) + (\text{biais } \hat{\theta})^2 ;$$

On calcule alors le rapport  $\frac{MSE(\hat{\theta}_1)}{MSE(\hat{\theta}_2)}$  :

Si  $\frac{MSE(\hat{\theta}_1)}{MSE(\hat{\theta}_2)} > 1$  alors l'estimateur  $\hat{\theta}_2$  est plus efficace que l'estimateur  $\hat{\theta}_1$  ;

Si  $\frac{MSE(\hat{\theta}_1)}{MSE(\hat{\theta}_2)} < 1$  alors l'estimateur  $\hat{\theta}_1$  est plus efficace que l'estimateur  $\hat{\theta}_2$ .

Il existe plusieurs méthodes d'estimation paramétrique dont on étudiera : la méthode du maximum de vraisemblance et la méthode des moments. Nous donnerons un exemple pour chaque méthode en guise d'illustration.

### III.1.1 Méthode du maximum de vraisemblance :

La vraisemblance est le moyen permettant d'identifier la distribution de laquelle dérive l'échantillon sur lequel on travaille et les paramètres inconnus à l'aide de données. Elle concerne le fait que tous les individus  $x_i$  de l'échantillon tiré proviennent de la même population étudiée.

La méthode du maximum de vraisemblance permet de sélectionner la distribution rendant maximale la vraisemblance de l'échantillon dont les individus sont tous tirés simultanément.

Soit  $X$  une variable aléatoire de densité de probabilité  $f(x, \theta)$  dont l'expression analytique est connue mais la valeur numérique de l'un des paramètres ou de tous les paramètres, en l'occurrence  $\theta$ , est inconnue.

On se propose donc l'inférence du paramètre  $\theta$  numériquement et ce, à partir de la construction d'une expression analytique fonction des réalisations de la variable aléatoire  $X$  dans un échantillon de taille  $n$ .

On définit la fonction de vraisemblance  $L$  (notation universelle) comme suit :

$$L(x_1, x_2, \dots, x_n, \theta) = f(x_1, \theta) * f(x_2, \theta) * \dots * f(x_n, \theta) = \prod_{i=1}^n f(x_i, \theta) ;$$

$\theta$  est un vecteur représentant les paramètres qu'on cherche. Il peut être unidimensionnel, si la fonction de densité de probabilité dépend uniquement d'un seul paramètre, ou multidimensionnel, si la fonction densité de probabilité dépend de plus d'un paramètre.

Le problème de tout estimateur est qu'il se calcule à partir d'un échantillon, alors qu'on cherche les paramètres de toute une population.

On cherche alors la valeur du paramètre  $\theta$ , pour laquelle la vraisemblance atteint son maximum tout en espérant que cette valeur soit proche de celle ayant engendré réellement l'échantillon.

C'est donc une méthode d'estimation et par suite un problème d'optimisation. Mathématiquement, il suffit de dériver la fonction  $L$  pour obtenir la valeur en question.

Pour simplifier les calculs, il est préférable de passer au logarithme de  $L$ . On parlera alors de log-vraisemblance notée log- $L$ .

Notons que puisque la fonction logarithmique est monotone et croissante, alors les fonctions vraisemblance et log-vraisemblance atteindront le même maximum. La valeur recherchée de  $\theta$  est donc la même dans les deux cas.

La fonction log-L est définie par :

$$L(x, \theta) = \prod_{i=1}^n f(x_i, \theta) \Rightarrow \log-L = \sum_{i=1}^n \log f(x_i, \theta)$$

La fonction L atteint un extrema  $\Leftrightarrow \frac{\partial L(x, \theta)}{\partial \theta} = 0$  (idem pour L-log)

Seulement, il n'est pas toujours facile de résoudre cette équation analytiquement. De plus, cette dernière indique seulement l'existence d'un extrema ou un point d'inflexion. Or l'extrema recherché est un maximum et l'équation reflétant le plus ce fait est :

$$\frac{\partial^2 L}{\partial \theta^2} < 0 ;$$

Le deuxième problème est que ce maximum peut ne pas être unique. Il faudra donc sélectionner la plus grande valeur de maximum parmi celles obtenues.

Notons que quand la taille de l'échantillon tend vers l'infini, la probabilité de l'unicité du maximum tend vers 1.

Exemple :

Dans cet exemple, nous allons détailler l'estimation des paramètres de la loi normale par la méthode du maximum de vraisemblance :

Comme vu dans le chapitre précédent, la densité de probabilité de la loi Normale s'exprime par :

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2} \left(\frac{x-\mu}{\sigma}\right)^2\right\} ; \text{ Avec : } \begin{cases} \mu: \text{moyenne} \\ \sigma: \text{écart - type} \end{cases}$$

Recherche de la fonction vraisemblance :

$$L = \prod_{j=1}^n f(x_j, \mu, \sigma^2) \Rightarrow$$

$$L = \prod_{j=1}^n \frac{1}{\sigma\sqrt{2\pi}} \exp \left\{ -\frac{1}{2} \left( \frac{x_j - \mu}{\sigma} \right)^2 \right\} \Rightarrow$$

$$L = \frac{n}{\sigma\sqrt{2\pi}} \left[ e^{-\frac{1}{2} \frac{\sum (x_j - \mu)^2}{\sigma^2}} \right] \Rightarrow$$

$$L = \left( \frac{1}{2\pi\sigma^2} \right)^{\frac{n}{2}} \left[ e^{-\frac{1}{2} \frac{\sum (x_j - \mu)^2}{\sigma^2}} \right] \Rightarrow$$

$$W = \ln L = \frac{-n}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum (x_j - \mu)^2 \Rightarrow$$

$$W = \frac{-1}{2} \left[ n \ln(2\pi\sigma^2) + \frac{1}{\sigma^2} \sum (x_j - \mu)^2 \right]$$

Nous devons, à présent, chercher à maximiser la fonction W représentant la Log-vraisemblance.

Pour ce faire, on annule les dérivées par rapport aux paramètres à estimer à savoir :  $\mu$  et  $\sigma$ . On aura donc :

$$\frac{\partial W}{\partial \mu} = \frac{1}{\sigma^2} \sum (x_j - \mu) = 0$$

$$\frac{\partial W}{\partial \sigma^2} = \frac{-n}{2\sigma^2} + \frac{2}{\sigma^4} \sum (x_j - \mu)^2 = 0$$

$$D'où : \sum (x_j - \mu) = 0 \Rightarrow$$

$$-n\sigma^2 + \sum (x_j - \mu)^2 = 0$$

Donc :

$$\sum x_j - n\mu = 0 \Rightarrow \hat{\mu} = \frac{1}{n} \sum x_j$$

$$\widehat{\sigma^2} = \frac{1}{n} \sum (x_j - \mu)^2$$

Notons que l'estimateur de la variance est un estimateur biaisé car :

$E(\widehat{\sigma^2}) - \sigma^2 \neq 0$ . Mais il convient toutefois de signaler, qu'il est asymptotiquement sans biais. C'est pourquoi, on préconise d'utiliser cette formule pour un échantillon dont la taille est supérieure ou égale à 30.

Pour s'affranchir de ce biais, on appliquera plutôt la formule suivante :

$$\widehat{\sigma^2} = \frac{1}{n-1} \sum (x_j - \mu)^2$$

Cet estimateur est sans biais. Il peut être donc utilisé, quelle que soit la taille de l'échantillon.

### III.1.2 Méthode des moments :

Cette méthode consiste à estimer les paramètres d'une distribution en égalisant les moments théoriques de celle-ci et les moments empiriques qui lui sont associés. Cette égalisation est justifiée par la loi des grands nombres qui permet la quasi-égalité de l'espérance mathématique et la moyenne empirique.

La résolution d'un système d'équations est donc à prévoir où le nombre d'équations est égal au nombre des paramètres à estimer.

Exemple :

Dans cet exemple, nous allons détailler l'estimation des paramètres de la loi gamma à deux paramètres par la méthode des moments :

La fonction densité de probabilité pour la loi de Gamma à deux paramètres s'exprime par :

$$f(x, t, \lambda) = \frac{\lambda e^{-\lambda x}}{\Gamma(t)} (\lambda x)^{t-1} ; \Gamma(t) \text{ représentant la fonction gamma.}$$

Les moments théoriques de cette distribution ont pour expression :

$$\begin{cases} E(X) = \frac{t}{\lambda} \\ V(X) = \frac{t}{\lambda^2} \end{cases} \Rightarrow \begin{cases} t = \lambda E(X) \\ V(X) = E(X^2) - E^2(X) = \frac{t}{\lambda^2} \Rightarrow V(X) = \frac{\lambda E(X)}{\lambda^2} \Rightarrow \end{cases}$$

$$\begin{cases} \lambda = \frac{E(X)}{E(X^2) - E^2(X)} \\ t = \frac{E^2(X)}{E(X^2) - E^2(X)} \end{cases} \Rightarrow \begin{cases} \hat{\lambda} = \frac{\frac{1}{n} \sum xi}{(\frac{1}{n} \sum xi^2 - (\frac{1}{n} \sum xi)^2)} \\ \hat{t} = \frac{(\frac{1}{n} \sum xi)^2}{(\frac{1}{n} \sum xi^2 - (\frac{1}{n} \sum xi)^2)} \end{cases}$$



Le test de l'adéquation d'une loi statistique constitue l'une des étapes les plus importantes dans une analyse statistique de données. Nous allons alors aborder ce point.

### III.2 Tests d'adéquation :

Le test d'adéquation consiste à vérifier le degré de vraisemblance de l'hypothèse selon laquelle les données analysées suivent la distribution testée. Pour ce faire, on définit deux types d'hypothèses appelées hypothèse nulle et hypothèse alternative.

L'hypothèse nulle constitue l'hypothèse selon laquelle la statistique de test choisie est vraie. L'hypothèse alternative, quant à elle, elle représente la négation de l'hypothèse nulle.

Selon les propriétés mathématiques du test, celui-ci peut être paramétrique ou non paramétrique.

Les tests paramétriques consistent à tester une hypothèse portant sur les paramètres de la distribution étudiée. Il est donc utilisé pour démontrer l'existence d'un effet comme la corrélation entre deux paramètres ou la différence entre caractéristiques statistiques.

Quant aux tests non paramétriques, contrairement aux tests paramétriques, ils servent à tester l'hypothèse selon laquelle la distribution théorique choisie s'ajuste à la distribution empirique étudiée. Ils ne font donc pas d'hypothèses particulières sur les paramètres de cette dernière.

Plusieurs tests de ce type peuvent être distingués dont nous citerons pour notre cas : le test du khi-deux, le test de Kolmogorov Smirnov, et le test d'Anderson-Darling.

Détaillons chacun des tests ci-dessus cités.

#### III.2.1 Test du Khi-deux :

Le test du khi-deux est un test asymptotique car il converge toujours vers une distribution khi-deux pour une grande taille N de l'échantillon.

On montre par le théorème de Pearson, que la distribution inconnue tend vers une Khi-deux quand la taille de l'échantillon tend vers l'infini.

La statistique de test du khi-deux Q s'écrit :

$$Q = \sum \frac{(O_i - E_i)^2}{E_i} ; \text{ avec : } \begin{cases} O_i : \text{effectifs observés} \\ E_i : \text{effectifs attendus} \end{cases}$$

Le test du Khi-deux concerne à l'origine les variables discrètes mais il est aussi applicable pour des variables continues et ce en discrétisant la distribution de probabilité continue  $F(x)$  en un nombre fini  $k$  de probabilités.

Mais, le choix de  $k$  est problématique car si  $k$  est petit, les tranches seront grandes, d'où la perte de l'information suite au regroupement car le test ne distinguera pas bien entre les différentes distributions d'où la faiblesse de sa puissance. Dans le cas contraire, le nombre d'observations par tranche sera très réduit ce qui rend le test imprécis.

La distribution asymptotique de la statistique du test khi-deux sera khi-deux  $(k-p-1)$  avec  $p$  représentant le nombre de degrés de liberté de la loi testée. En d'autres termes, l'estimation de  $p$  paramètres de la distribution testée aura pour conséquence la perte de  $p$  degrés khi-2 de liberté de la distribution asymptotique.

### III.2.2 Test de Kolmogorov-Smirnov KS:

Le test de Kolmogorov-Smirnov est un test libre non-paramétrique qui a pour hypothèse nulle le test de l'adéquation de la loi testée.

La statistique de ce test est :  $D_n = \text{Sup}\{|Fn(x) - F(x)|\}$ . Le calcul de cette statistique se réduit donc à un tri de l'échantillon, à l'évaluation des fréquences empirique et théorique sur les points de l'échantillon et à la recherche du maximum des différences calculées.

L'hypothèse nulle est acceptée si :

$(D_n < d)$  ; avec:  $d$  valeur lue à partir de la table de KS selon la taille de l'échantillon  $n$ .

### III.2.3 Test d'Anderson-Darling:

Ce test vient corriger la faiblesse de celui de Cramer-Von Mises à savoir que la différence entre les fonctions empirique et théorique tend vers zéro quand  $x$  tend vers l'infini. La sensibilité de la statistique s'avère donc faible par rapport aux positions réelles des queues de distribution.

Le test d'Anderson-Darling est un test d'adéquation qui représente une modification du test de Kolmogorov-Smirnov et donne plus de poids aux données que celui de Cramer-Von mises en faisant intervenir la fonction de pondération dont la formule est la suivante :

$$\frac{1}{F(x)(1 - F(x))}$$

Le test K-S est une distribution libre (free distribution) dans le sens où les valeurs critiques ne sont pas prises en considération dans la distribution testée.

Le test d'Anderson-Darling utilise la distribution spécifiée dans le calcul des valeurs critiques. Ce qui avantage ce test car il prend en considération les valeurs extrêmes. Mais l'inconvénient est qu'il faudra calculer les valeurs critiques pour chaque distribution.

Il existe actuellement des tables où figurent les valeurs critiques pour certaines lois.

La statistique du test d'Anderson-Darling est :

$$A^2 = -N - S ; \text{ Avec :}$$

$$S = \sum_{i=1}^N \frac{2i-1}{N} (\ln(F(X_i)) + \ln(1 - F(X_{n+1-i})))$$

Où :  $F(X_i)$  est la fonction de distribution cumulative de la distribution considérée

$X_i$  : les réalisations ordonnées selon l'ordre croissant ou décroissant.

Les valeurs pour certaines lois ont été tabulées. L'hypothèse nulle selon laquelle la loi testée s'ajuste à l'échantillon est rejetée si la statistique du test A est supérieure à la valeur critique.

Les tests que nous allons détailler ci-après doivent être effectués préalablement à toute analyse et ce, afin d'assurer sa significativité du point de vue statistique.

Ainsi, pour valider théoriquement une étude statistique, il faut que les propriétés suivantes soient satisfaites : l'aléa de la variable étudiée, l'indépendance, l'homogénéité et la stationnarité de la chronique étudiée.

Pour vérifier ces propriétés, plusieurs tests doivent être effectués en l'occurrence :

### III.3 Tests d'échantillonnage :

#### III.3.1 Tests d'homogénéité de Wilcoxon et de Pettitt :

Connu aussi sous l'appellation « test de Mann-Whitney », le test de Wilcoxon est un test non paramétrique libre basé sur la comparaison des moyennes. Il sert à vérifier l'appartenance de tous les individus de l'échantillon tiré à la même population. Lors de l'application de ce test, on suppose que les termes de la série sont indépendants et identiquement distribués.

L'idée de base de ce test, est que si l'on mélange les valeurs des deux sous-échantillons formés et que l'on ordonne toutes les valeurs par ordre croissant, on doit obtenir un mélange homogène auquel cas, les rangs moyens de chacun des sous-échantillons seront sensiblement égaux si l'hypothèse nulle à savoir que les médianes sont égales, est vérifiée.

La statistique de ce test se calcule comme suit :

Nous devons tout d'abord former deux sous-échantillons de tailles  $n_1$  et  $(n_2=n-n_1)$  à partir de la série étudiée puis attribuer un rang à toutes les valeurs classées par ordre croissant. Pour les valeurs appariées, on leur attribuera un rang égal à la moyenne de leurs rangs.

On calcule les sommes  $S_1$  et  $S_2$  des rangs des éléments constituant chaque sous-échantillon puis les variables suivantes :

$$\begin{cases} U_a = (n_1 n_2) + \frac{n_1(n_1+1)}{2} - (S_1) \\ U_b = (n_1 n_2) + \frac{n_2(n_2+1)}{2} - (S_2) \end{cases} ;$$

$$\begin{cases} U_1 = \frac{U_a - \vartheta}{\sigma} \\ U_2 = \frac{U_b - \vartheta}{\sigma} \end{cases} ; \text{Où: } \begin{cases} \vartheta = n_1 * n_2 / 2 \\ \sigma = \sqrt{n_1 * n_2 * (n_1 + n_2 + 1) / 12} \end{cases}$$

$$U = \min \{U_1, U_2\}$$

Si  $U < u(1-\alpha/2) \Rightarrow$  l'hypothèse d'homogénéité est vérifiée.

Le test de Pettitt est une modification du test de Mann-Whitney dont l'hypothèse nulle est l'absence de rupture dans la série.

La mise en œuvre du test suppose que pour tout instant  $t$  variant de 1 à  $N$ , les séries  $(x_i)$ , pour  $(i=1,t)$  et  $(i=t+1,N)$  appartiennent à la même population.

Soit  $D_{ij} = \text{sgn}(x_i - x_j)$  avec  $(\text{sgn})$  représentant la fonction signe.

On considère la variable  $U_{t,n}$  telle que :

$$U_{t,n} = \sum_{i=1}^t \sum_{j=t+1}^N D_{ij} ;$$

Soit  $K_n$  la variable définie par le maximum en valeur absolue de  $U_{t,n}$  pour  $t=1, N-1$ .

Si  $k$  désigne la valeur de  $K_n$  prise sur la série étudiée, sous l'hypothèse nulle, la probabilité de dépassement de la valeur  $k$  est donnée approximativement par :

$$P(K_n > k) = 2 \exp\left(\frac{-6k^2}{(N^3 + N^2)}\right);$$

Pour un risque de première espèce  $\alpha$  donné, si :

$P(K_n > k) < \alpha$ , alors l'hypothèse nulle est rejetée. La série sera donc non homogène.

### III.3.2 Test de stationnarité KPSS :

Prosaïquement, la stationnarité désigne la constance des caractéristiques statistiques d'une chronique au cours du temps. Par conséquent, celles-ci devraient rester invariantes quel que soit l'instant  $t$  du processus.

Dans l'étude des crues, la notion de stationnarité s'avère très importante sur le plan pratique, car une augmentation dans le régime des événements extrêmes aurait des conséquences bien plus graves qu'une simple évolution dans les conditions moyennes.

Notons que la propriété de stationnarité est définie pour un processus et non pour une série chronologique. Cependant, par abus de langage, on parle souvent de série stationnaire pour parler d'une série possédant les propriétés d'un processus stationnaire.

Avant toute analyse statistique, on a besoin de vérifier si la série temporelle étudiée contient une tendance. Pour ce faire, nous devons analyser la stationnarité de la chronique en tendance. Il existe alors plusieurs tests statistiques entre autres le test KPSS.

Le test KPSS (Kwiatkowski-Phillips-Schmidt-Shin) est un test non paramétrique dont la statistique s'exprime par :

$$K = \frac{\sum St^2}{n^2 \sigma^2};$$

Avec :

$$\left\{ \begin{array}{l} \sigma: \text{variance de long terme estimée par l'estimateur non paramétrique de Newey – West} \\ St: \text{vecteur des résidus de la régression} \\ n: \text{taille de l'échantillon} \end{array} \right.$$

### III.3.3 Test d'indépendance de Wald-Wolfowitz :

L'indépendance des valeurs est une autre condition qui vient s'ajouter à celles citées précédemment pour minimiser les erreurs d'échantillonnage. Cette notion signifie que toutes les réalisations de la variable aléatoire ont les mêmes chances d'apparaître, il n'existe donc pas d'auto-corrélation.

Ce test est utilisé pour vérifier l'indépendance des valeurs constituant l'échantillon.

La statistique du test est la suivante :

$$R = \sum_{i=1}^{n-1} x_i x_{i+1} + x_1 x_n$$

$R$  suit une loi Normale de moyenne  $\bar{R}$  et de variance  $V(R)$  sachant que :

$$\left\{ \begin{array}{l} \bar{R} = \frac{S_1^2 - S_2}{n-1} \\ S_r = nm_r = \sum_{i=1}^n (x_i - \bar{x})^r \\ V(R) = \frac{S_1^2 - S_4}{n-1} - \bar{R}^2 + \frac{S_1^4 - 4S_1^2 S_2 + 4S_1 S_3 + S_2^2 - 2S_4}{(n-1)(n-2)} \end{array} \right.$$

On calcule par la suite, la variable réduite :  $u = \frac{R - \bar{R}}{\sqrt{V(R)}}$

Pour un niveau de signification  $\alpha$  % l'hypothèse d'indépendance est acceptée si :  $|u| \leq u_{1-\alpha/2}$  avec :  $u_{1-\alpha/2}$  lue à partir de la table de Laplace-Gauss.

### III.3.4 Test des horsains de Grubbs et Beck :

Ce test a pour objectif de signaler la présence de valeurs singulières appelées horsains, et a pour statistique la variable suivante :

Après avoir calculé la statistique de ce test, on calculera les valeurs maximale et minimale :

$$\begin{cases} Xh = e^{(\sigma u + \bar{x})} \\ Xl = e^{(-\sigma u + \bar{x})} \end{cases}$$

Les valeurs de la chronique étudiée doivent appartenir à l'intervalle  $[Xl, Xh]$ .

Dans le cas où l'une des valeurs n'appartient pas à cet intervalle, on conclura que celle-ci est une valeur horsain.

Dans le chapitre suivant, nous allons présenter l'approche de minimisation des critères pénalisés dont nous avons parlé précédemment.

## Chapitre IV : Méthodes bayésiennes : critères d'information AIC-BIC

Les méthodes bayésiennes consistent à introduire des connaissances a priori sur les paramètres des échantillons étudiés provenant d'analyses statistiques antérieures. Cela se traduit par des probabilités d'occurrence pour les différentes valeurs possibles du paramètre, appelées probabilité a priori.

Pour un modèle  $M_i$ , la distribution a priori du paramètre  $\theta_i$ , est notée :

$$P(\theta_i/M_i).$$

Celle-ci est ensuite actualisée à l'aide des données observées en appliquant la formule de Bayes. Ainsi, on aura la loi a posteriori du paramètre à estimer.

Il existe une littérature très fournie sur ces méthodes qui est en train de se développer actuellement avec l'apparition d'outils complexes de probabilités.

L'une des notions ayant pour base ce genre de méthodes est celle de la minimisation des critères d'information pénalisés à savoir les critères AIC et BIC.

Ces derniers permettent la comparaison entre plusieurs modèles pour en sélectionner le plus adéquat et ce, en affectant une fonction de pénalité à la fonction objectif (économique) afin de pénaliser les points ne respectant pas les contraintes (égalité ou inégalité que doit satisfaire la solution du problème).

Ces critères se placent dans un contexte bayésien : le paramètre  $\theta_i$  et les modèles candidats  $M_i$  sont vus comme des variables aléatoires. La distribution a priori est notée  $P(M_i)$ .

Les critères d'information sont basés sur le maximum de vraisemblance en pénalisant la vraisemblance des paramètres estimés par cette méthode qui reflètent la qualité de l'ajustement. Ils sont considérés comme de bons indicateurs du meilleur compromis entre l'ajustement et la parcimonie du modèle.

Les critères s'écrivent en général sous la forme :

$$-2 \log L(\xi) + 2|\xi|f(n) ;$$

- $f(n)$  est la fonction pénalisation dépendant de la taille de l'échantillon  $n$ ; celle-ci change selon le critère étudié ;



- $L(\xi)$  est la vraisemblance des paramètres à estimer et  $\xi$  est le nombre de paramètres.

### IV.1 Le critère d'information d'Akaike AIC (1973) :

L'AIC est donc la pénalisation de la log-vraisemblance par deux fois le nombre de paramètres  $k$ . Il a donc pour formule :

**AIC =  $-2 \ln L(\theta) + 2k$**  ;  $L$  représentant la vraisemblance maximisée et  $k$  représentant le nombre de paramètres du modèle et  $F(n)$  étant égale à l'unité par définition.

### IV.2 Le critère d'information Bayésien BIC (Schwarz, 1978) :

Comme le critère AIC, le critère BIC cherche le modèle maximisant la probabilité a posteriori  $P(M_i|X)$ , ainsi on pourra écrire alors que :

$$M_{\text{BIC}} = \operatorname{argmax} P(M_i|X)$$

En ce sens, le critère BIC cherche à sélectionner le modèle le plus vraisemblable au vu des données.

La formule de base utilisée est celle de Bayes qui s'écrit comme suit :

$$P(M_i|X) = \frac{P(X|M_i)P(M_i)}{P(X)}$$

Les distributions a priori pour les modèles candidats  $M_i$  sont supposées égales (non informatives) :  $P(M_1)=P(M_2)= \dots = P(M_n)$ . Ainsi, aucun modèle n'est privilégié. La recherche du meilleur modèle ne nécessite que le calcul de la probabilité  $P(X|M_i)$ .

Cette dernière est appelée vraisemblance intégrée pour le modèle  $M_i$ . Elle a pour expression l'intégrale suivante :

$$\int P(X, \theta_i|M_i) d\theta_i = \int g(X, \theta_i)P(\theta_i|M_i)d\theta_i;$$

$g$  représentant la vraisemblance correspondant au modèle  $M_i$  de paramètre  $\theta_i$ .

Son calcul exact est donc rarement possible vu la complexité de la forme mathématique de la vraisemblance de la distribution considérée. On utilise l'approximation de Laplace en guise d'estimation de cette intégrale.

Considérons la fonction :

$$L_n(\theta_i) = \frac{g(\theta_i)}{n} = \frac{1}{n} \sum_{k=1}^n \log(g(M_i(X_k, \theta_i))) + \frac{\log(P(\theta_i|M_i))}{n}$$

$$\text{On pose : } \begin{cases} \theta_i^* = \operatorname{argmax} L_n(\theta_i) \\ A(\theta_i^*) = -\left(\frac{\partial^2 L_n(\theta_i)}{\partial \theta_i^j \partial \theta_i^l}\right) \Big|_{\theta_i = \theta_i^*} \end{cases} ;$$

$A(\theta_i^*)$  représentant l'opposé de la matrice hessienne des dérivées secondes partielles  
 $\theta_i^j$  est la jème composante du vecteur des paramètres  $\theta_i$

On aura alors :

$$P(X|M_i) = e^{\log(g(\theta_i^*))} (P(\theta_i^*|M_i)) \left(\frac{2\pi}{n}\right)^{\frac{K_i}{2}} |A(\theta_i^*)|^{-0.5} + O(n^{-1})$$

En passant au logarithme de l'expression obtenue, on aura :

$$\log P(X|M_i) = \log(g(X, \theta_i^*)) + \log(P(\theta_i^*|M_i)) + \frac{K_i}{2} \log(2\pi) - 0.5 \log(|A(\theta_i^*)|) + O(n^{-1}) ;$$

O fait référence à la notation de Landau.

A cette étape, on fait les approximations suivantes :

$$\begin{cases} \hat{\theta}_i = \operatorname{argmax} \left( \frac{g(X, \theta_i^*)}{n} \right) : \text{estimateur par maximum de vraisemblance} \\ A(\widehat{\theta}_i^*) = -E \left( \frac{\partial^2 \log(X_l, \theta_i)}{\partial \theta_i^j \partial \theta_i^l} \right) \Big|_{\theta_i = \theta_i^*} : \text{matrice d'information de Fisher} \end{cases}$$

On remplace alors ces deux approximations dans l'équation précédente :

$$\log P(X|M_i) = \log(g(X, \hat{\theta}_i)) - \frac{K_i}{2} \log(n) + \log(P(\hat{\theta}_i|M_i)) + \frac{K_i}{2} \log(2\pi) - 0.5 \log(|A(\widehat{\theta}_i^*)|) + O(n^{-0.5})$$

Avec :

- $\{\log(g(X, \hat{\theta}_i)) - \frac{K_i}{2} \log(n)\}$  tend vers  $-\infty$  avec  $n$
- $\log(P(\hat{\theta}_i|M_i)) + \frac{K_i}{2} \log(2\pi) - 0.5 \log(|A(\widehat{\theta}_i^*)|)$  reste borné :  $O(1)$

En négligeant les termes  $O(1)$  et  $O(n^{-0.5})$ , on aura :

$$\log P(X|Mi) = \log \left( g(X, \hat{\theta}_i) \right) - \frac{K_i}{2} \log(n)$$

Pour le critère BIC, celui-ci est égal, pour un modèle  $M_i$ , à :  $-2\log P(X|Mi)$

D'où :

$$\mathbf{BIC} = \mathbf{-2 \ln L(\boldsymbol{\theta}) + k \ln(n)} ;$$

Le BIC est donc la pénalisation de la log-vraisemblance par le nombre de paramètres  $k$  multiplié par la taille de l'échantillon. Ainsi, plus le nombre d'observations  $n$  augmente, plus la pénalité augmente. Cette pénalisation dépasse 2 (pour  $n > 7$ ) et donc le BIC a tendance à sélectionner des modèles plus petits que l'AIC.

Ce critère a été proposé pour le choix du modèle pour des échantillons de taille élevée pour lesquels l'AIC a tendance à choisir des modèles à nombre élevé de paramètres, tandis que le BIC aboutit à des modèles plus parcimonieux.

Il assure aussi la convergence vers le quasi-vrai modèle, même si celui-ci est emboîté dans un modèle plus général. C'est cette propriété que l'on appelle « consistance pour la dimension ».

Le modèle à retenir est celui qui montre les critères calculés AIC ou BIC les plus faibles.

Après avoir exposé les principales notions mathématiques et statistiques régissant le choix du modèle, on exposera dans le chapitre suivant la partie application de notre travail.

## Chapitre V : Choix du modèle adéquat pour des séries de précipitations maximales

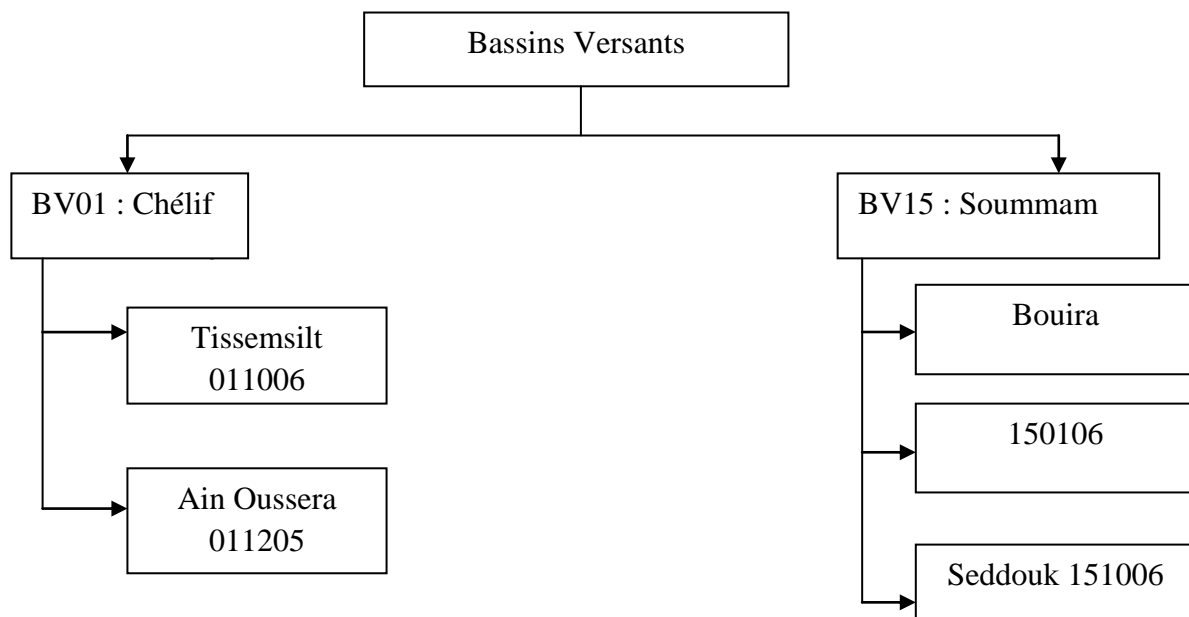
### V.1 Etude des temps de retour et des fréquences empiriques :

Dans un premier temps, nous allons analyser, les valeurs des probabilités et des temps de retour données par les différentes formules de fréquence empirique mentionnées dans le chapitre précédent, en l'occurrence : Hazen, Weibull, Chegodayev, Cunnane, Gringorten et Tukey et ce, pour les différentes valeurs constituant l'échantillon considéré.

Pour appliquer cette méthode, les étapes à suivre sont les suivantes :

- Classement de l'échantillon par ordre croissant ;
- Attribution d'un rang à chacune des valeurs ainsi classées ;
- Calcul des fréquences empiriques correspondantes ;
- Tracé des courbes :  $x = f$  (Probabilité au non dépassement) et temps de retour  $x = f(T)$ .

Les chroniques étudiées représentent les précipitations maximales journalières, comme le montre le schéma ci-après :



Ainsi, Nous présenterons les résultats obtenus pour une chronique de précipitations journalières maximales représentant chacun des deux bassins versants étudiés, en l'occurrence : la station d'Ain Oussera et la station de Bouira.

Les résultats obtenus pour le reste des séries temporelles étudiées, seront présentés en annexe.

Pour cette première partie, le même travail a été fait pour les différentes séries temporelles étudiées. Etant donné que nous sommes arrivés aux mêmes conclusions pour ces dernières, nous ne présentons que les résultats obtenus pour la station d'Ain Oussera.

- Résultats de la station d'Ain Oussera :

Avant d'énoncer les résultats obtenus pour la chronique de cette station, nous allons présenter les résultats du prétraitement des données.

La chronique étudiée pour cette station représente les pluies journalières maximales pour une période d'observation de 60 ans dont la valeur maximale est égale à (81 mm) et la valeur minimale est égale à (12.3 mm).

Les caractéristiques statistiques de l'échantillon étudié sont :

<p><b>Moyenne=30.39 ;</b></p> <p><b>Ecart-type=11.84 ;</b></p> <p><b>Coefficient d'asymétrie Cs=1.28;</b></p> <p><b>Coefficient d'aplatissement</b> <b>Ck=4.16;</b></p> <p><b>Coefficient de variation Cv=0.39.</b></p>
---

Les paramètres caractéristiques de l'échantillon composé par les logarithmes des valeurs sont les suivants :

<p><b>Moyenne=3.34;</b></p> <p><b>Ecart-type=0.39;</b></p> <p><b>Coefficient d'asymétrie Cs= -0.16;</b></p> <p><b>Coefficient d'aplatissement</b></p> <p><b>Ck= -0.06;</b></p> <p><b>Coefficient de variation Cv=0.12.</b></p>
--

Nous devons maintenant vérifier, de prime abord, les tests de stationnarité, d'homogénéité et d'indépendance.

Les résultats de ces tests sont résumés dans les tableaux suivants :

Test de Wold-Wolfowitz	Test de Pettitt	Test KPSS
Statistique : $ u =0.738$	P-value : $p=0.985$	Statistique : $ K =0.095$
Décision : $u < 1,96 \Rightarrow$ <b>Indépendance vérifiée à un niveau de confiance de 95%</b>	Décision : $p > 0.05 \Rightarrow$ <b>Homogénéité vérifiée à un niveau de confiance de 95%</b>	Décision: $K < 0.146 \Rightarrow$ <b>Stationnarité vérifiée à un niveau de confiance de 95%</b>

Les résultats des tests de cette chronique étant positifs, nous pouvons alors l'analyser statistiquement.

Le résultat donné par le test de Grubbs et Beck est résumé dans le tableau ci-après :

Hypothèse nulle	Statistique du test $K_n$	Valeur maximale (Xh)	Valeur minimale (Xl)
H0 : il existe des valeurs horsains : n'appartenant pas à l'intervalle [Xl , Xh]	2.84	85.30 mm	9.34 mm

Les queues droite et gauche de la chronique étudiée étant respectivement égales à 12.3 mm et 81 mm, on conclut que l'échantillon ne contient pas de valeurs horsains.

Tous les tests ci-dessus mentionnés étant vérifiés, l'analyse statistique de la chronique est significative.

Les résultats obtenus pour la méthode détaillée ci-dessus, sont résumés dans les graphes ci-après :

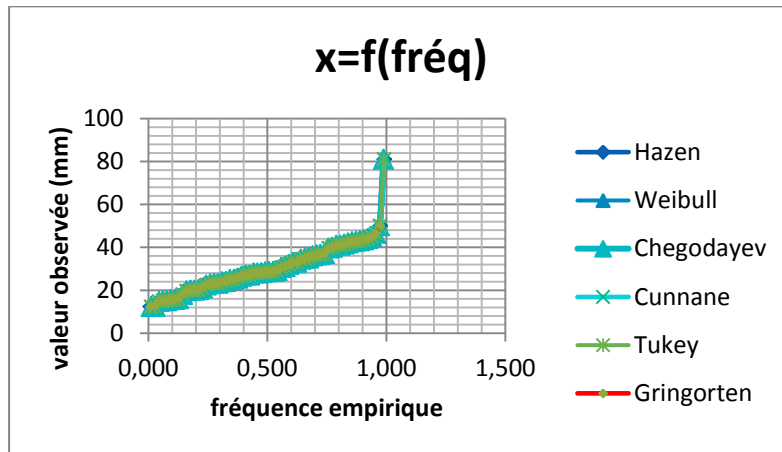


Figure V.1: courbe des valeurs observées en fonction de la fréquence au dépassement

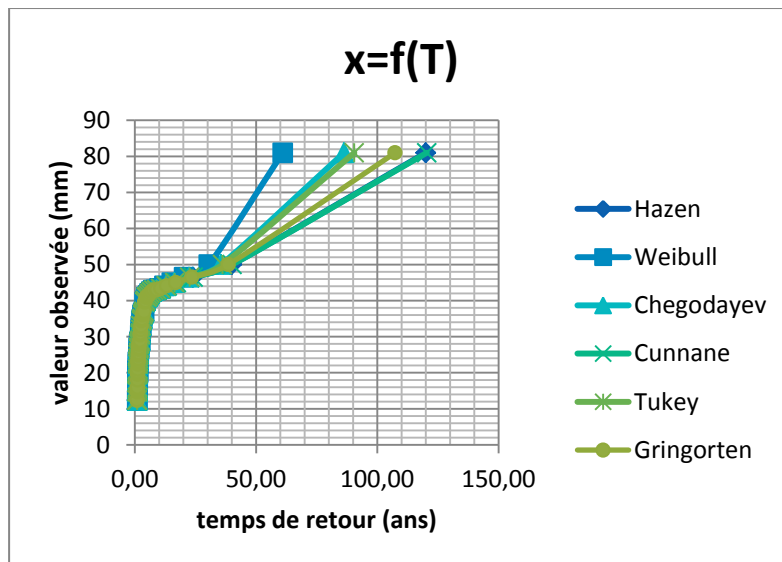


Figure V.2: Courbe des valeurs observées en fonction du temps de retour

On remarque à travers ces graphes que pour :

**P<sub>j</sub>=81 mm**, et **P<sub>j</sub>=50 mm**, les temps de retour obtenus diffèrent selon la loi utilisée et ce, malgré que les fréquences empiriques au non dépassement soient relativement proches :

<b>P<sub>j</sub>=81 mm</b>	Hazen	Weibull	Chegodayev	Cunnane	Tukey	Gringorten
T (ans)	120.00	61.00	86.29	86.00	90.50	107.36
FD	0.008	0.016	0.012	0.012	0.011	0.009
FND	0.992	0.984	0.988	0.992	0.989	0.991
<b>P<sub>j</sub>=50 mm</b>						
T (ans)	40.00	30.50	35.53	35.41	36.20	38.54
FD	0.025	0.033	0.028	0.028	0.028	0.026
FND	0.975	0.967	0.972	0.975	0.972	0.974

C'est la même constatation pour les plus grandes valeurs de la chronique étudiée, sauf pour les valeurs autres qu'extrêmes où, le temps de retour n'est pas très différent.

Nous concluons alors que pour les probabilités au dépassement, la différence n'est pas importante tandis qu'elle l'est pour les périodes de retour des queues droites.

En conséquence, le problème se pose au niveau des valeurs les plus extrêmes où les résultats sont différents selon la formule utilisée.

Cela étant, la formule de fréquence la plus préconisée actuellement est celle de Cunnane. On optera alors pour cette dernière. Cela dit, il convient de signaler que chacune des formules empiriques de fréquence précitées, devrait correspondre à une certaine loi de probabilité.

Le choix de la fréquence empirique pour une distribution donnée doit être tel que le point expérimental se situe en moyenne sur la courbe théorique. On peut écrire alors que :

$$F_i = F(E(x_r)) = P(X < E(x_r));$$
 r représentant le rang de la valeur observée x.



Pour une loi à deux paramètres, il est plus simple de faire le changement de variable suivant :

$$U = \frac{x - x_0}{a}$$

On aura alors :

$$(E(xr)) = P(X < E(xr)) = P\left(\frac{x - x_0}{a} < \frac{E(xr) - x_0}{a}\right) = P(U < E(Ur))$$

Pour la loi exponentielle, on montre que :

$$(E(xr)) = P(X < E(xr)) = \sum_{j=1}^r \frac{1}{N + 1 - j}$$

$$\text{D'où : } F(E(xr)) = 1 - e^{-E(Ur)} ;$$

Ainsi, en calculant la fréquence théorique de la loi exponentielle pour les différents rangs  $r$ , nous pourrons la comparer aux valeurs de la fréquence empirique pour chaque formule, pour en déduire celle qui s'approche le mieux de la fréquence théorique.

En effectuant les calculs pour les différentes séries que nous avons étudiées, on peut conclure que la forme la plus proche de la fréquence théorique de la loi exponentielle est celle obtenue par la formule de Cunnane et ce, principalement au niveau des valeurs extrêmes de la chronique, les valeurs sont pratiquement identiques.

### V.2 Choix du modèle adéquat pour des séries de précipitations maximales :

Comme mentionné précédemment, les séries que nous allons présenter sont celles des précipitations maximales journalières de la station de Bouira et d'Ain Oussera.

Dans cette partie, nous allons nous intéresser au choix de la distribution statistique qui s'ajuste le mieux aux séries temporelles étudiées.

Pour ce faire, on doit en premier lieu vérifier les tests d'homogénéité, d'indépendance, de stationnarité et des valeurs horsains pour assurer la validité de l'analyse statistique à effectuer.

En deuxième lieu, on ajustera visuellement les différentes lois considérées à savoir : loi Normale, la loi Log-Normale, la loi Gumbel, la loi Pearson 3, la loi Log-Pearson 3, la loi Weibull, la loi exponentielle et enfin la loi GEV.

En troisième lieu, on procédera à la vérification des ajustements par les tests d'adéquation suivants : khi-deux, Kolmogorov-Smirnov et Anderson-Darling.

En quatrième lieu, on appliquera la méthode d'harmonisation des quantiles et des temps de retour. L'approche consiste à calculer les quantiles pour une période de retour donnée ainsi que le temps de retour de la valeur extrême droite des séries considérées et ce, pour les différentes distributions choisies, en l'occurrence : la loi normale, la loi log-Normale, la loi Gumbel, la loi exponentielle, la loi Pearson 3, la loi log-Pearson 3, et la loi Weibull et la loi GEV.

Après avoir calculé les quantiles, on devra comparer les valeurs obtenues avec les valeurs observées tout en tenant compte de la période d'observation. On devrait donc choisir la loi obéissant le mieux à ces conditions. Cela dépend alors de la capacité du modèle considéré à décrire les événements effectivement observés.

En définitive, nous calculerons les critères d'information d'Akaike et Bayésien (AIC et BIC) afin de choisir la meilleure loi de probabilité.

Nous allons présenter dans ce qui suit, les résultats obtenus en appliquant les différentes méthodes ci-dessus explicitées. Après quoi, nous énoncerons le meilleur modèle choisi selon les critères AIC et BIC.

**V.2.1 Etude de la Station d'Ain Oussera :**

Nous allons tout d'abord rappeler les résultats des tests d'échantillonnage et les paramètres statistiques énoncés précédemment (cf. supra V.1).

Les caractéristiques statistiques de l'échantillon étudié sont :

<p><b>Moyenne=30.39 ;</b></p> <p><b>Ecart-type=11.84 ;</b></p> <p><b>Coefficient d'asymétrie Cs=1.28;</b></p> <p><b>Coefficient d'aplatissement</b> <b>Ck=4.16;</b></p> <p><b>Coefficient de variation Cv=0.39.</b></p>
---

Les paramètres caractéristiques de l'échantillon composé par les logarithmes des valeurs sont les suivants :

<p><b>Moyenne=3.34;</b></p> <p><b>Ecart-type=0.39;</b></p> <p><b>Coefficient d'asymétrie Cs= -0.16;</b></p> <p><b>Coefficient d'aplatissement</b> <b>Ck= -0.06;</b></p> <p><b>Coefficient de variation Cv=0.12.</b></p>
---

Les résultats des tests de stationnarité, d'homogénéité et d'indépendance sont résumés dans les tableaux suivants :

<b>Test de Wold-Wolfowitz</b>	<b>Test de Pettitt</b>	<b>Test KPSS</b>
Statistique : <b> u =0.738</b>	P-value : <b>p=0.985</b>	Statistique : <b> K =0.095</b>
Décision : <b>u&lt;1,96 =&gt;</b> <b>Indépendance vérifiée à un niveau de confiance de 95%</b>	Décision : <b>p&gt;0.05=&gt;</b> <b>Homogénéité vérifiée à un niveau de confiance de 95%</b>	Décision: <b>K&lt;0.146 =&gt;</b> <b>Stationnarité vérifiée à un niveau de confiance de 95%</b>

Les tests pour cette chronique étant positifs, nous pouvons alors l'analyser statistiquement.

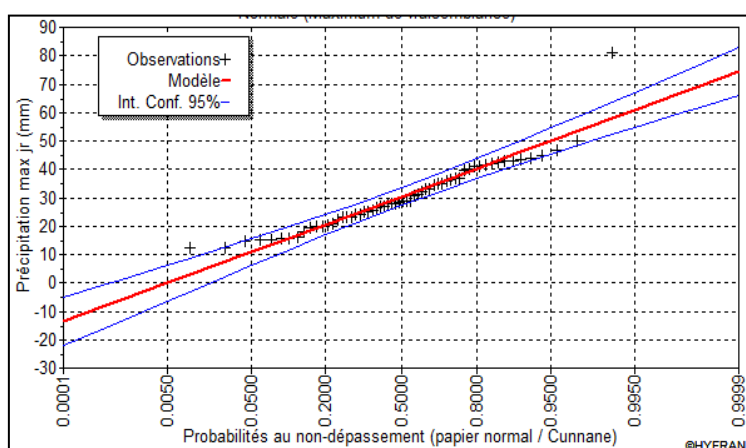
Le résultat donné par le test de Grubbs et Beck est résumé dans le tableau ci-après :

Hypothèse nulle	Statistique du test $K_n$	Valeur maximale ( $X_h$ )	Valeur minimale ( $X_l$ )
$H_0$ : il existe des valeurs horsains : n'appartenant pas à l'intervalle $[X_l, X_h]$	2.84	85.30 mm	9.34 mm

Les queues droite et gauche de la chronique étudiée étant respectivement égales à 12.3 mm et 81 mm, on conclut que l'échantillon ne contient pas de valeurs horsains.

Tous les tests ci-dessus mentionnés étant vérifiés, l'analyse statistique de la chronique est significative.

L'ajustement visuel mettant en évidence les intervalles de confiance pour les différentes lois de probabilité traitées peut être représenté par les graphiques suivants :



**Figure V.3 : Ajustement graphique de la loi Normale**

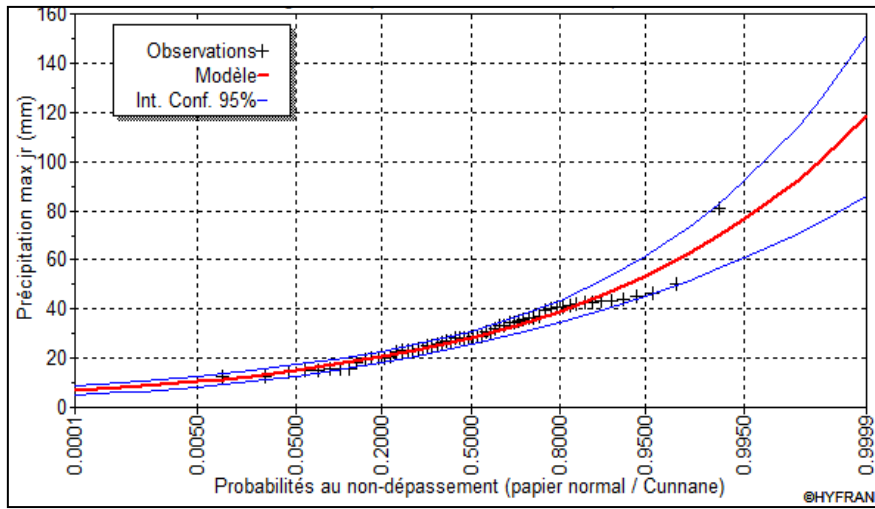


Figure V.4: Ajustement graphique de la loi Log-Normale

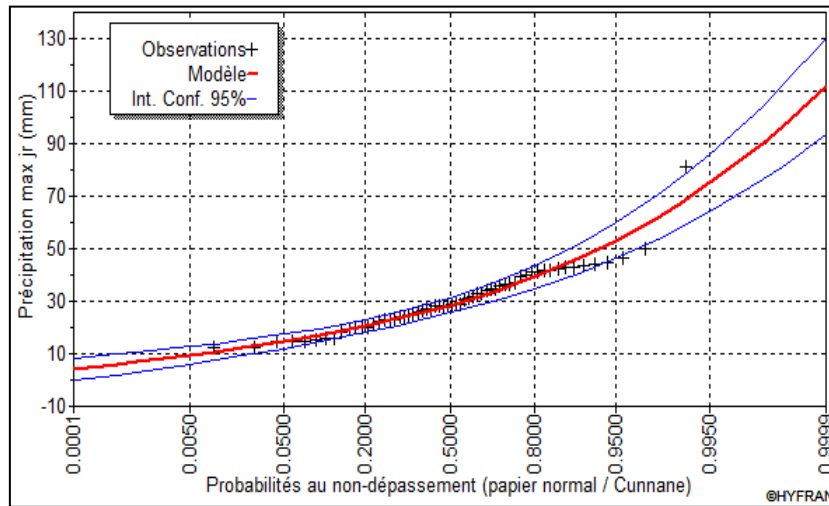


Figure V.5: Ajustement graphique de la loi Gumbel

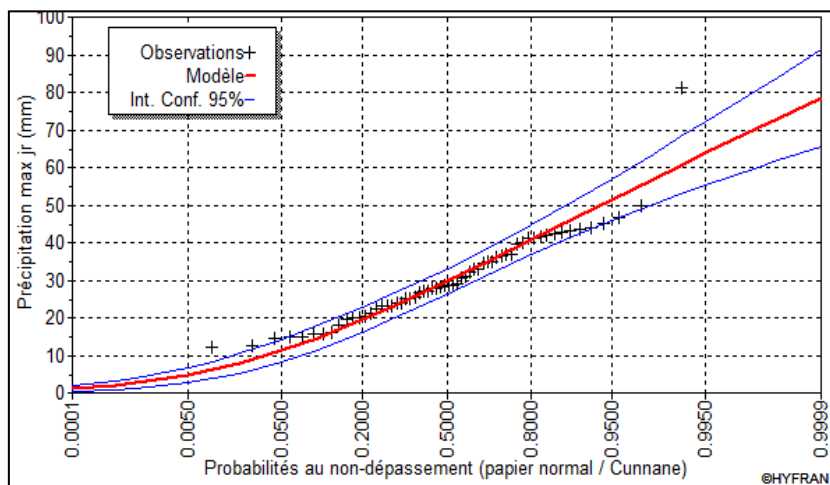


Figure V.6: Ajustement graphique de la loi Weibull

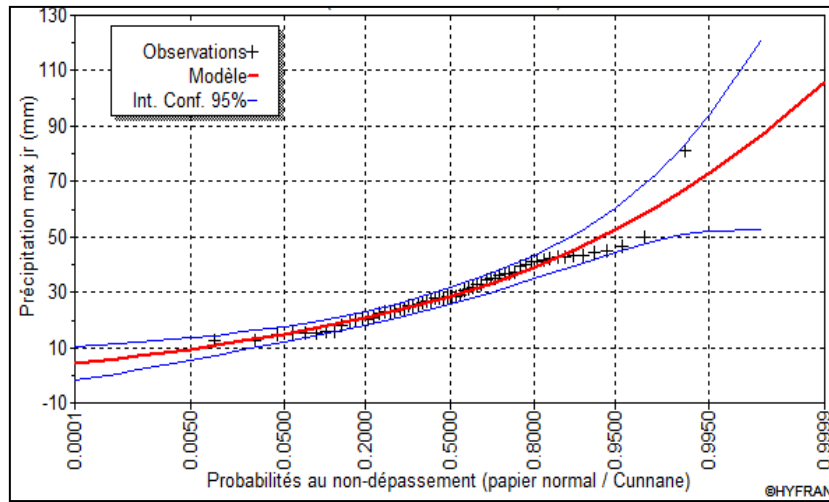


Figure V.7: Ajustement graphique de la loi GEV

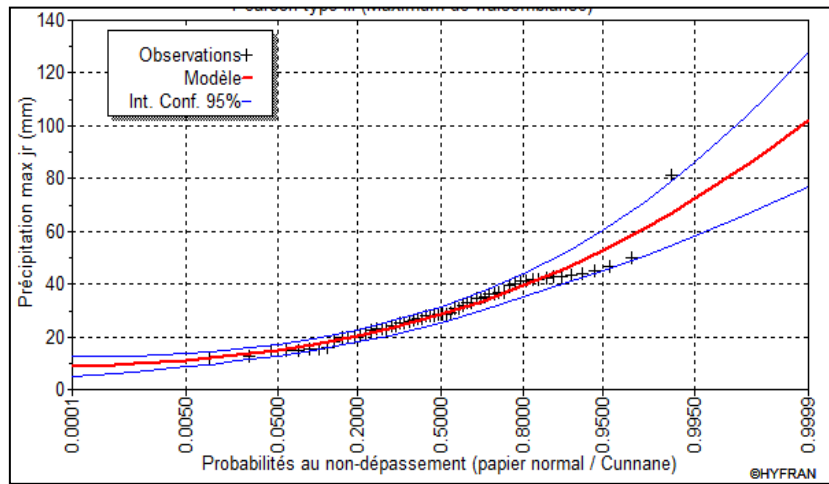


Figure V.8: Ajustement graphique de la loi Pearson 3

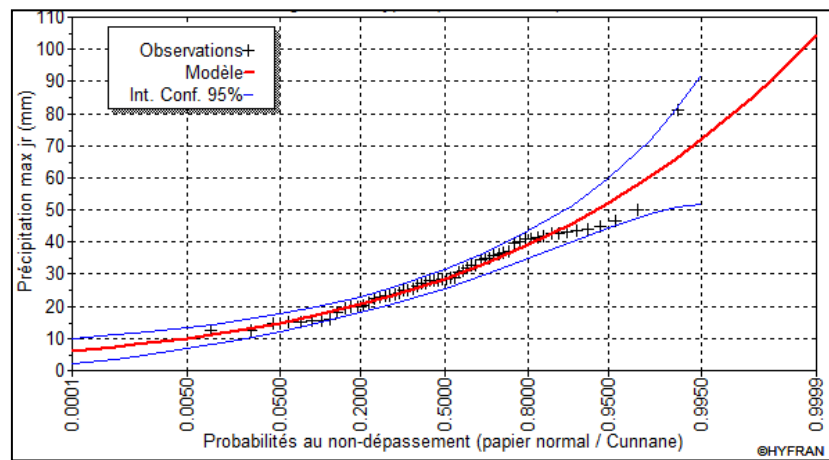


Figure V.9: Ajustement graphique de la loi Log-Pearson 3

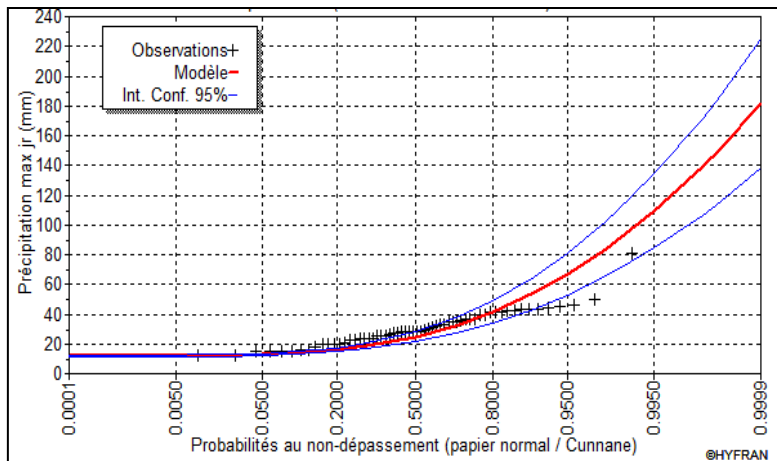


Figure V.10 : Ajustement graphique de la loi exponentielle

A travers ces différents graphiques d'ajustement visuel, il est clair que la loi exponentielle est à rejeter, celle-ci s'ajustant très mal à l'échantillon avec plusieurs points en dehors des courbes enveloppes des intervalles de confiance, en plus du fait que comparées au reste des lois, les points ne s'alignent pas sur la courbe du modèle.

Pour les lois restantes, il est difficile de choisir parmi elles, celle correspondant le mieux à l'échantillon ; en effet, elles semblent toutes, peu ou prou, s'ajuster à ce dernier.

Nous allons maintenant présenter un autre type d'ajustement visuel qui est le diagramme Quantile-Quantile.

### Diagrammes Quantile-Quantile :

Ci-après sont présentées les figures représentant les diagrammes quantile-quantile obtenus pour les différentes distributions étudiées :

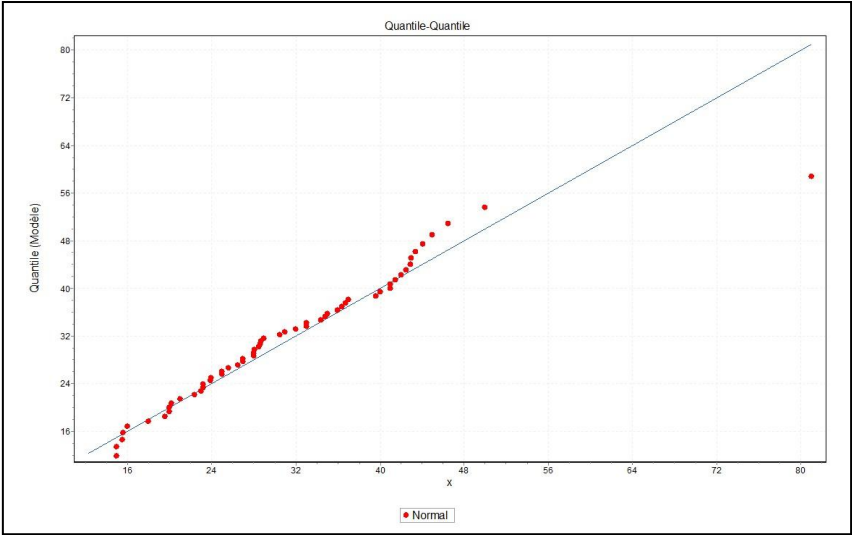


Figure V.11: Q-Q plot pour la loi Normale

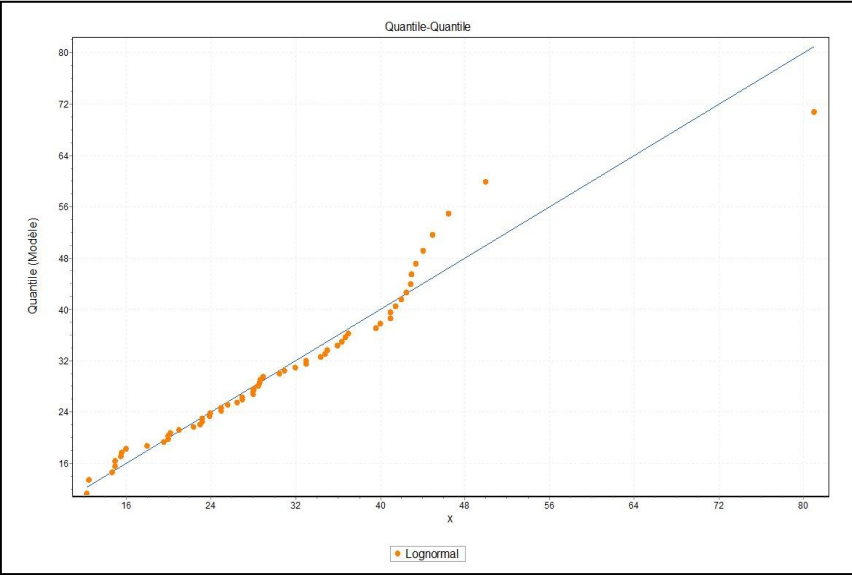


Figure V.12: Q-Q plot pour la loi Log-Normale



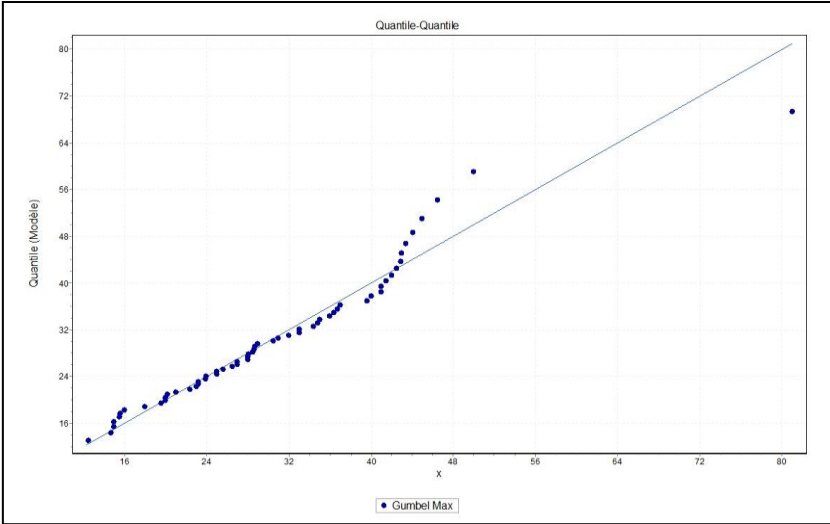


Figure V.13: Q-Q plot pour la loi Gumbel

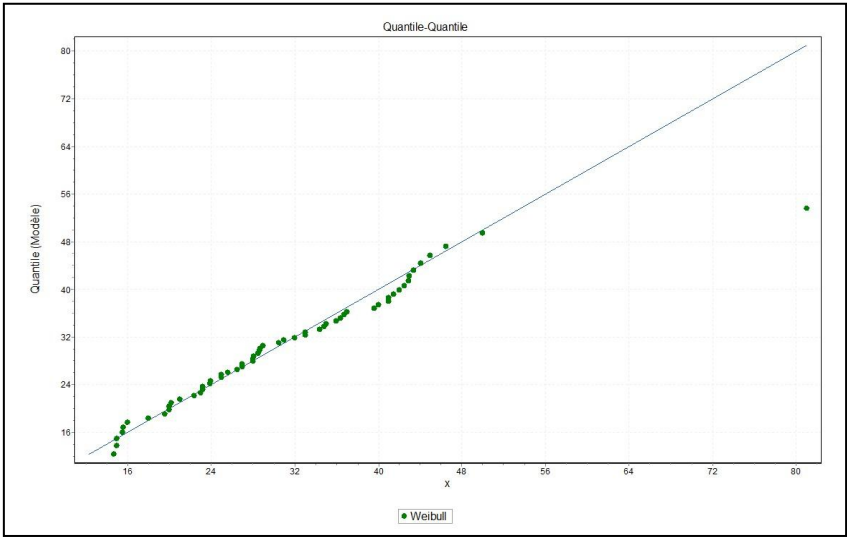


Figure V.14 : Q-Q plot pour la loi Weibull

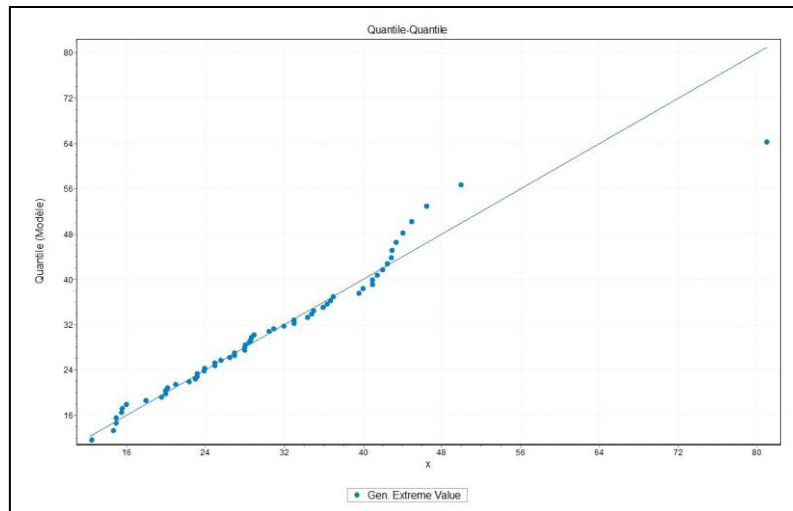


Figure V.15 : Q-Q plot pour la loi GEV

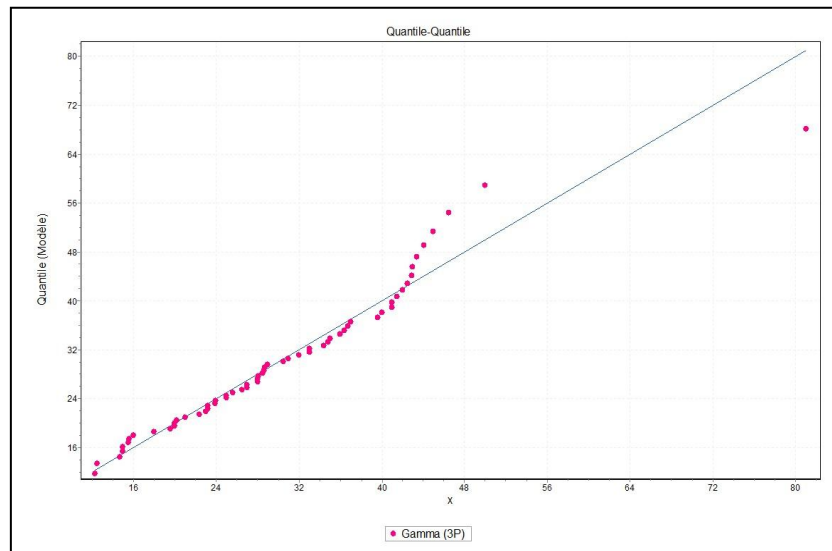


Figure V.16 : Q-Q plot pour la loi Pearson 3

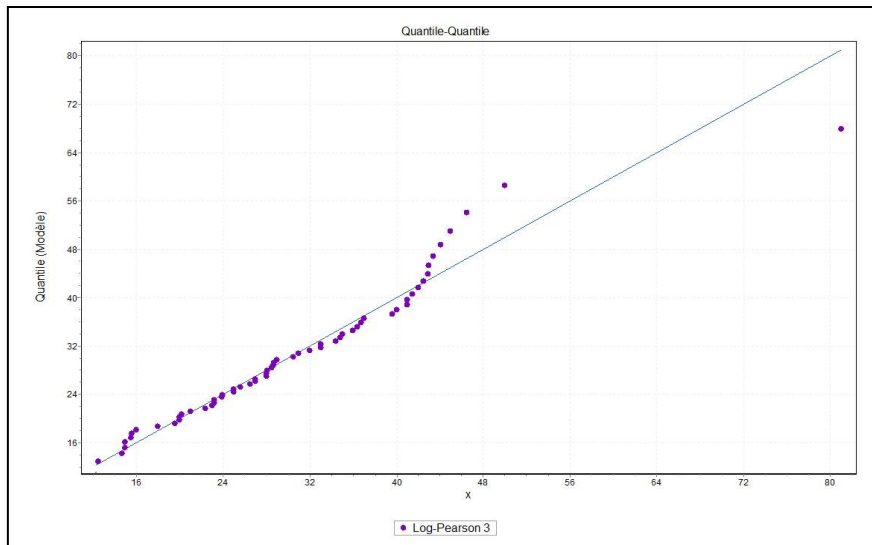


Figure V.17 : Q-Q plot pour la loi Log-Pearson 3

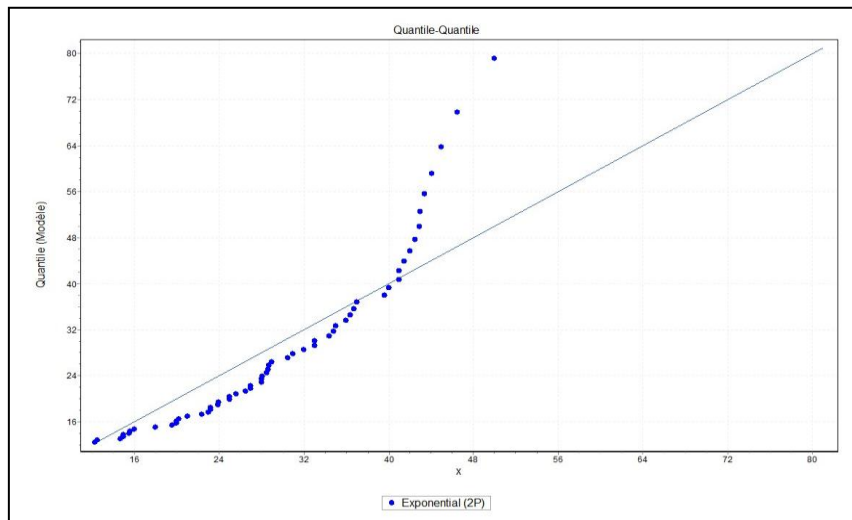


Figure V.18 : Q-Q plot pour la loi exponentielle

Nous remarquons à travers ces diagrammes que toutes les lois semblent montrer un nuage de points s'alignant relativement autour de la première bissectrice. Ce qui n'est pas le cas pour les lois Exponentielle et Normale.

Ainsi, la visualisation de ces graphiques nécessite-t-elle d'être étayée par une opération consistant à tester l'ajustement des différentes lois de probabilité citées précédemment, par les tests d'adéquation pris en compte en l'occurrence : khi-deux, K-S et Anderson-Darling.

Les résultats obtenus sont présentés dans le tableau suivant :

**Tableau V.1:** Tableau résumant les résultats des tests d'adéquation pour la chronique d'Ain Oussera.

Loi	Tests	Statistique du test : stat	Valeur critique: c	Décision
Normale	Khi-deux	<b>1.008</b>	11.07	<b>Stat&lt;c =&gt; ajustement accepté</b>
	K-S	<b>0.096</b>	0.172	<b>Stat&lt;c =&gt; ajustement accepté</b>
	A-D	<b>0.632</b>	2.502	<b>Stat&lt;c =&gt; ajustement accepté</b>
Log-Normale	Khi-deux	<b>1,156</b>	11.07	<b>Stat&lt;c =&gt; ajustement accepté</b>
	K-S	<b>0,065</b>	0.172	<b>Stat&lt;c =&gt; ajustement accepté</b>
	A-D	<b>0,495</b>	2.502	<b>Stat&lt;c =&gt; ajustement accepté</b>
Gumbel	Khi-deux	<b>2.395</b>	11.07	<b>Stat&lt;c =&gt; ajustement accepté</b>
	K-S	<b>0.064</b>	0.172	<b>Stat&lt;c =&gt; ajustement accepté</b>
	A-D	<b>0.457</b>	2.502	<b>Stat&lt;c =&gt; ajustement accepté</b>
Weibull	Khi-deux	<b>4.599</b>	11.07	<b>Stat&lt;c =&gt; ajustement accepté</b>
	K-S	<b>0.085</b>	0.172	<b>Stat&lt;c =&gt; ajustement accepté</b>
	A-D	<b>0.622</b>	2.502	<b>Stat&lt;c =&gt; ajustement accepté</b>
Expo	Khi-deux	<b>12.041</b>	9.488	<b>Stat&gt;c =&gt; ajustement rejeté</b>
	K-S	<b>0.196</b>	0.172	<b>Stat&gt;c =&gt; ajustement rejeté</b>
	A-D	<b>3.872</b>	2.502	<b>Stat&gt;c =&gt; ajustement rejeté</b>
GEV	Khi-deux	<b>2.000</b>	11.07	<b>Stat&lt;c =&gt; ajustement accepté</b>
	K-S	<b>0.059</b>	0.172	<b>Stat&lt;c =&gt; ajustement accepté</b>
	A-D	<b>0.355</b>	2.502	<b>Stat&lt;c =&gt; ajustement accepté</b>
Pearson 3	Khi-deux	<b>0.633</b>	11.07	<b>Stat&lt;c =&gt; ajustement accepté</b>
	K-S	<b>0.064</b>	0.172	<b>Stat&lt;c =&gt; ajustement accepté</b>
	A-D	<b>0.466</b>	2.502	<b>Stat&lt;c =&gt; ajustement accepté</b>
Log-Pearson3	Khi-deux	<b>2.322</b>	11.07	<b>Stat&lt;c =&gt; ajustement accepté</b>
	K-S	<b>0.061</b>	0.172	<b>Stat&lt;c =&gt; ajustement accepté</b>
	A-D	<b>0.409</b>	2.502	<b>Stat&lt;c =&gt; ajustement accepté</b>

On remarque à travers ce tableau que les trois tests utilisés donnent des résultats positifs quant à l'ajustement de toutes les lois de probabilité étudiées sauf pour la loi exponentielle.

A présent, nous allons appliquer la méthode basée sur l'harmonisation des quantiles et des temps de retour.

Les résultats obtenus pour la chronique étudiée sont les suivants :

Les quantiles calculés pour les différentes lois ainsi que les temps de retour de la plus grande valeur de la chronique étudiée sont résumés dans le tableau suivant :

**Tableau V.2 :** Tableau résumant les quantiles et les temps de retour calculés pour la chronique d'Ain Oussera

<b>Loi</b>	<b>Quantiles pour T= 60 ans (mm)</b>	<b>Temps de retour (ans) pour Pjmax=81 mm</b>
Normale	<b>55.6</b>	<b>3571.43</b>
Log-Normale	<b>64.3</b>	<b>357.14</b>
Gumbel	<b>63.6</b>	<b>376.69</b>
Pearson3	<b>62.3</b>	<b>68.41</b>
Log-Pearson3	<b>61.8</b>	<b>66.03</b>
Weibull	<b>57.9</b>	<b>54.55</b>
Exponentielle	<b>87.3</b>	<b>185.18</b>
GEV	<b>62.3</b>	<b>310.51</b>

Selon les résultats obtenus, comme le montre le tableau ci-dessus, tous les quantiles calculés pour une période de retour de 60ans ne correspondent pas à la plus grande valeur observée qui est de (81mm).

De plus, les temps de retour correspondant à la queue droite de la série sont différents de la période d'observation, à savoir 60 ans.

On en conclut que par la méthode d'harmonisation des quantiles, on ne peut choisir le modèle adéquat.

**V.2.2 Etude de la Station de Bouira :**

La chronique étudiée pour cette station représente les pluies journalières maximales pour une période d’observation de 53 ans.

Les caractéristiques statistiques de l’échantillon étudié sont :

<p><b>Moyenne=37.34;</b></p> <p><b>Ecart-type= 11.14 ;</b></p> <p><b>Coefficient d’asymétrie Cs=1.30 ;</b></p> <p><b>Coefficient d’aplatissement Ck=3.12 ;</b></p> <p><b>Coefficient de variation Cv=0.29.</b></p>
--

Les caractéristiques statistiques des logarithmes des valeurs de l’échantillon étudié sont :

<p><b>Moyenne=3.58 ;</b></p> <p><b>Ecart-type= 0.28 ;</b></p> <p><b>Coefficient d’asymétrie Cs= 0.21 ;</b></p> <p><b>Coefficient d’aplatissement Ck=0.45 ;</b></p> <p><b>Coefficient de variation Cv=0.079.</b></p>
---

Les tests de stationnarité, d’homogénéité et d’indépendance donnent les résultats suivants :

<b>Test de Wold-Wolfowitz</b>	<b>Test de Pettitt</b>	<b>Test KPPS</b>
Statistique : <b><math> u = 0.648</math></b>	P-value : <b><math>P = 0.202</math></b>	Statistique: <b><math> K =0.0287</math></b>
Décision : <b><math>u &lt; 1,96 \Rightarrow</math> Indépendance vérifiée à un niveau de confiance de 95%</b>	Décision : <b><math>P &gt; 0.05 \Rightarrow</math> Homogénéité vérifiée à un niveau de confiance de 95%</b>	Décision : <b><math>K &lt; 0.146 \Rightarrow</math> Stationnarité vérifiée à un niveau de confiance de 95%</b>

### Application du test de Grubbs et Beck

Le résultat obtenu pour le test est résumé dans le tableau suivant :

Hypothèse nulle	Statistique (Kn)	Valeur maximale (mm)	Valeur minimale (mm)
H0 :il existe des valeurs horsains : n'appartenant pas à l'intervalle [Xl , Xh]	2.79	78,36	16,42

Ce test indique l'incohérence de l'appartenance de la valeur 79.1mm à la chronique étudiée car étant supérieure à la valeur maximale. Elle est donc à retirer.

### Etude de la chronique sans la valeur horsain

La chronique étudiée pour cette station représente les pluies journalières maximales mais pour une période d'observation de 52 ans car on supprime la valeur horsain.

Les caractéristiques statistiques de l'échantillon étudié sont :

<p><b>Moyenne=36.54 ;</b></p> <p><b>Ecart-type=9.58;</b></p> <p><b>Coefficient d'asymétrie Cs=0.66 ;</b></p> <p><b>Coefficient d'aplatissement Ck=0.82 ;</b></p> <p><b>Coefficient de variation Cv=0.26.</b></p>
--

Les paramètres caractéristiques de l'échantillon composé par les logarithmes des valeurs sont les suivants :

<p><b>Moyenne=3.57 ;</b></p> <p><b>Ecart-type= 0.26 ;</b></p> <p><b>Coefficient d'asymétrie Cs= - 0.11 ;</b></p> <p><b>Coefficient d'aplatissement</b></p> <p><b>Ck= -0.07 ;</b></p> <p><b>Coefficient de variation Cv=0.07.</b></p>
--

Les tests de stationnarité, d'homogénéité et d'indépendance donnent les résultats suivants :

<b>Test de Wold-Wolfowitz</b>	<b>Test de Pettitt</b>	<b>Test KPSS</b>
Statistique : <b> u =0.118</b>	P-value : <b>P=0.145</b>	Statistique : <b> K =0.0491</b>
Décision : <b>u&lt;1,96 =&gt;</b> <b>Indépendance vérifiée à un niveau de confiance de 95%</b>	Décision : <b>P&gt;0.05 =&gt;</b> <b>Homogénéité vérifiée à un niveau de confiance de 95%</b>	Décision : <b>K&lt;0.146=&gt;</b> <b>Stationnarité vérifiée à un niveau de confiance de 95%</b>

Les résultats des tests de cette chronique étant positifs, nous pouvons alors l'analyser statistiquement.

Tous les tests ci-dessus mentionnés étant vérifiés, l'analyse statistique de la chronique est significative.

L'ajustement visuel mettant en évidence les intervalles de confiance pour les différentes lois de probabilité traitées peut être représenté par les graphiques suivants :



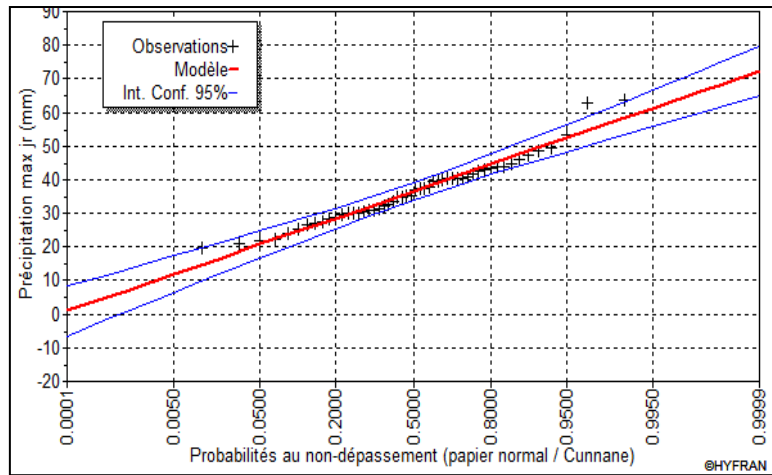


Figure V.19 : Ajustement graphique de la loi Normale

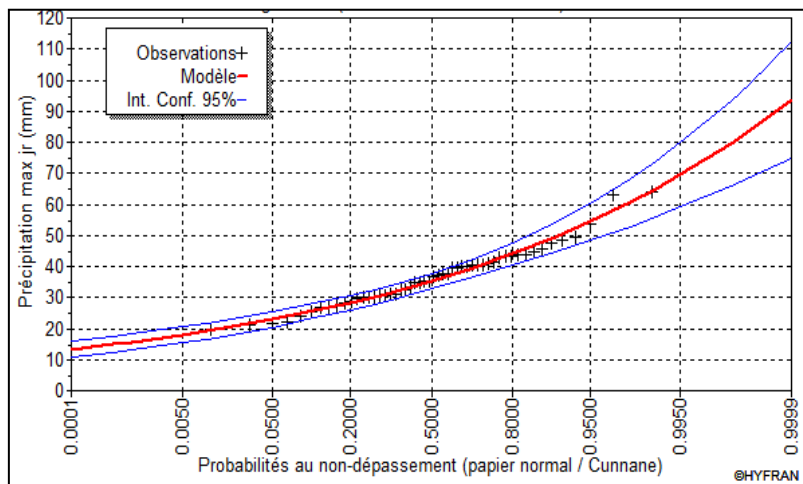


Figure V.20 : Ajustement graphique de la loi Log-Normale

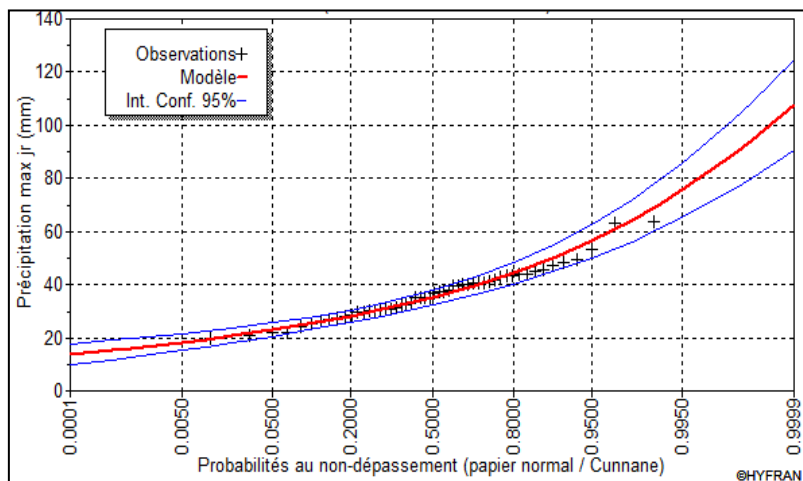


Figure V.21: Ajustement graphique de la loi Gumbel

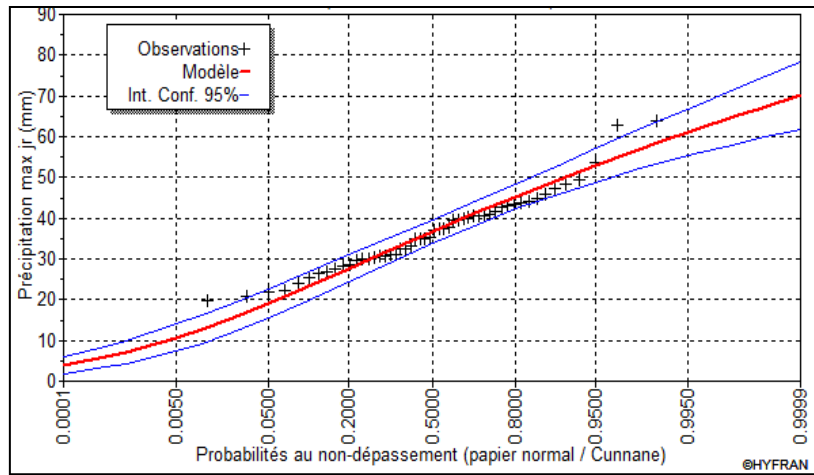


Figure V.22 : Ajustement graphique de la loi Weibull

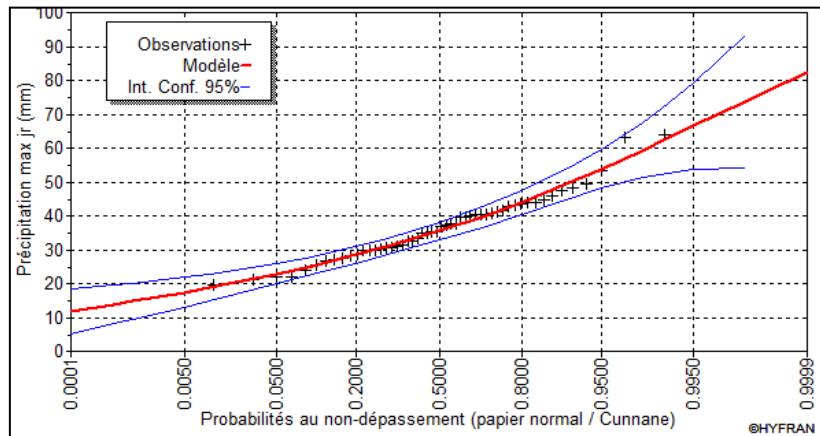


Figure V.23 : Ajustement graphique de la loi GEV

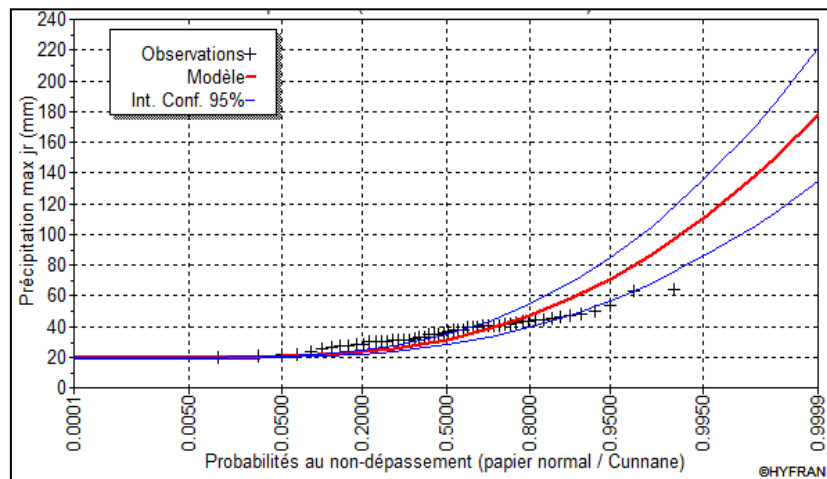


Figure V.24 : Ajustement graphique de la loi exponentielle

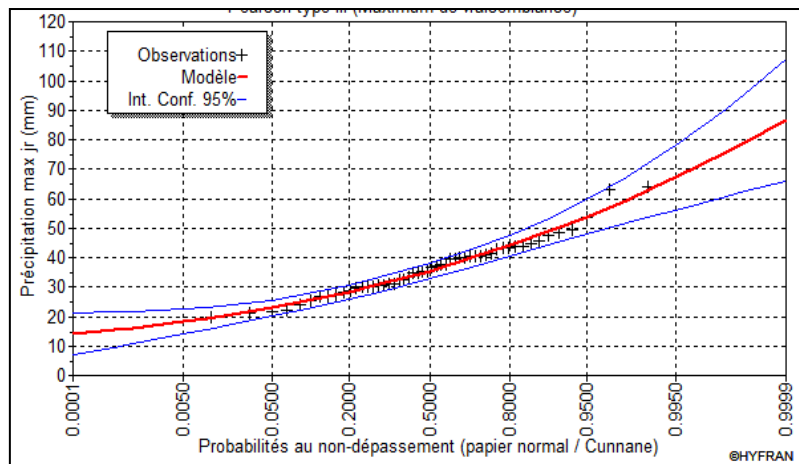


Figure V.25 : Ajustement graphique de la loi Pearson 3

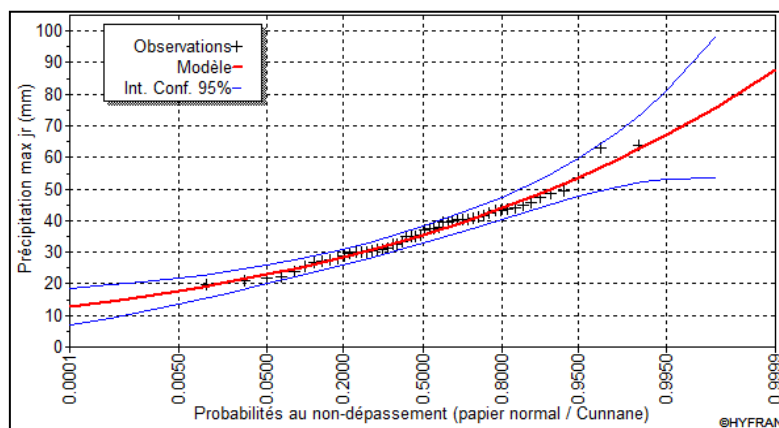


Figure V.26: Ajustement graphique de la loi Log-Pearson 3

On remarque à travers ces différents graphiques d'ajustement visuel que la loi exponentielle est à écarter car il existe certains points qui ne sont pas contenus dans son intervalle de confiance.

Pour les lois restantes, il est difficile de choisir parmi elles, celle correspondant le mieux à l'échantillon car elles semblent toutes s'ajuster relativement à ce dernier.

Présentons maintenant les graphiques Quantile-Quantile obtenus pour la chronique de Bouira.

### Diagrammes Quantile-Quantile :

Ci-après sont présentées les figures représentant les diagrammes quantile-quantile obtenus pour les différentes distributions étudiées :

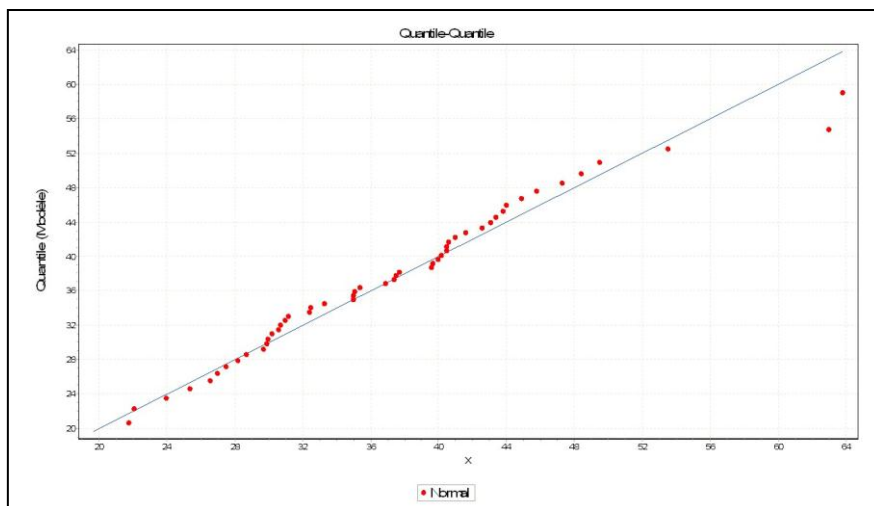


Figure V.27 : Q-Q plot pour la loi Normale

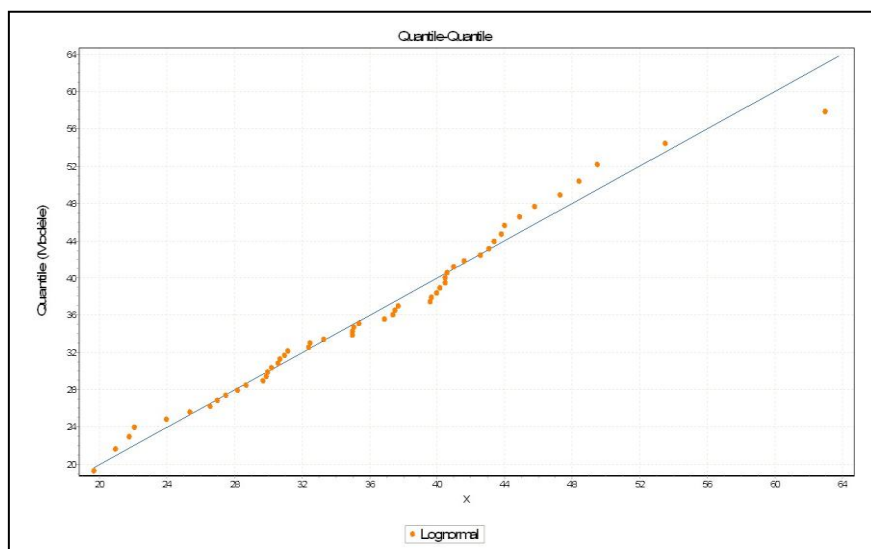


Figure V.28 : Q-Q plot pour la loi Log-Normale

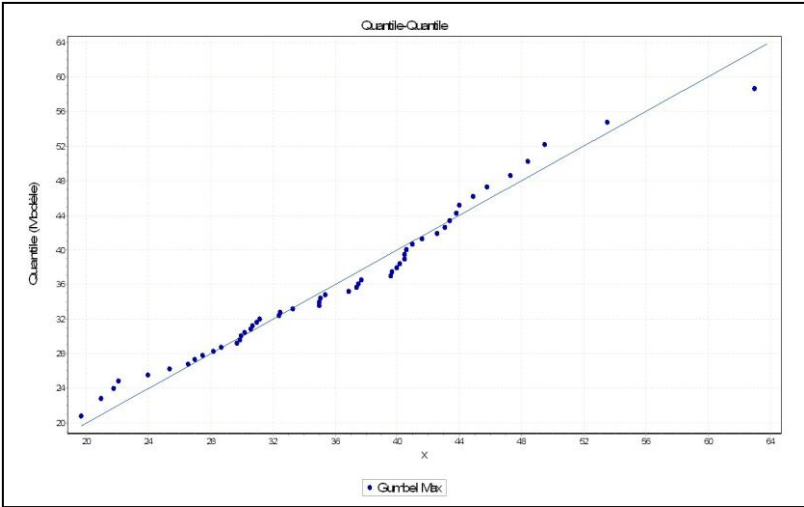


Figure V.29 : Q-Q plot pour la loi Gumbel

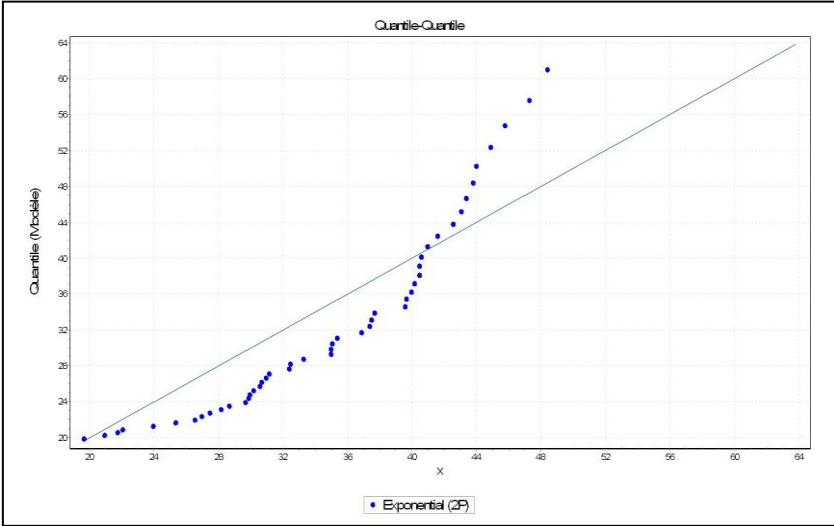


Figure V.30 : Q-Q plot pour la loi exponentielle

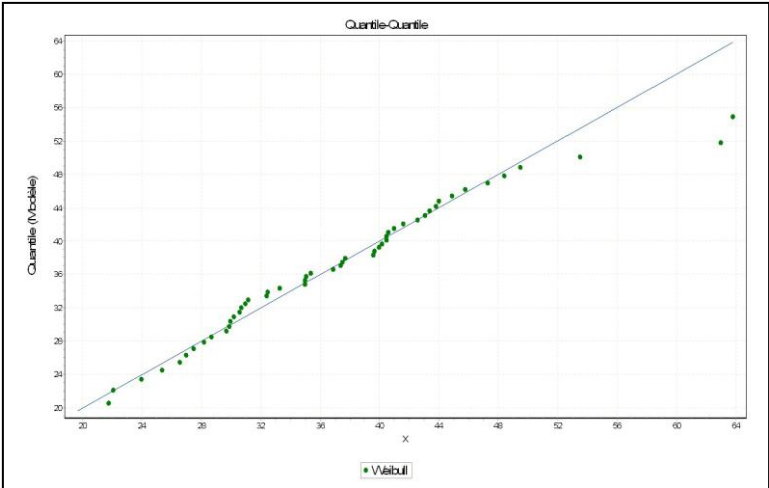


Figure V.31 : Q-Q plot pour la loi Weibull

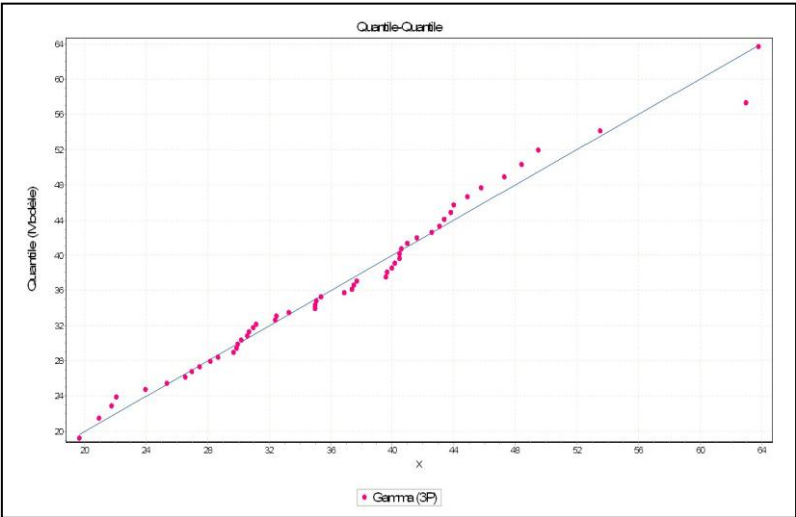


Figure V.32 : Q-Q plot pour la loi Pearson 3

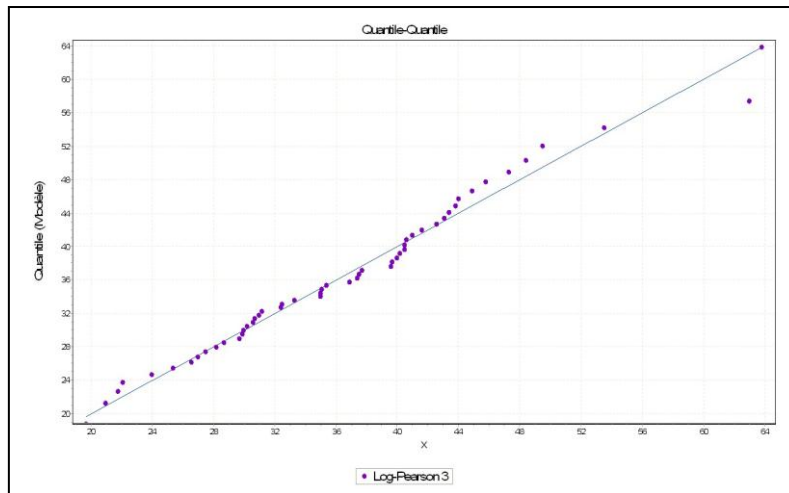


Figure V.33: Q-Q plot pour la loi Log-Pearson 3

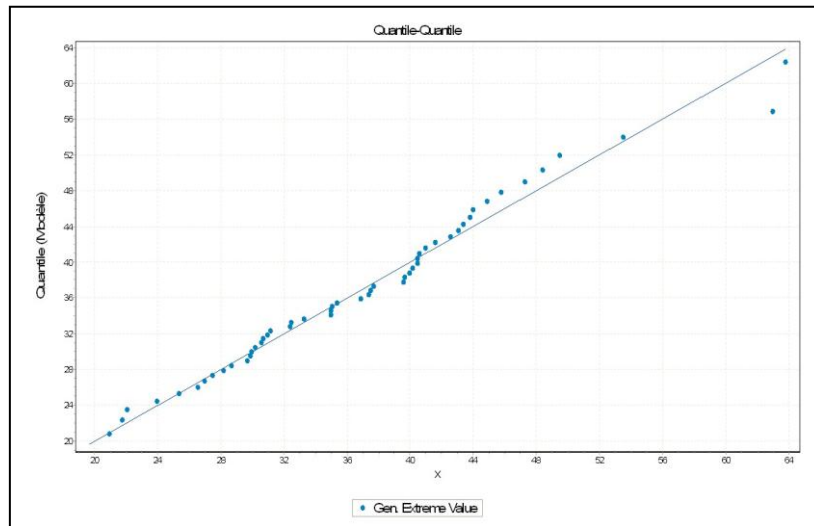


Figure V.34 : Q-Q plot pour la loi GEV

On remarque à travers ces diagrammes que toutes les lois et particulièrement la loi exponentielle s'ajustent mal à l'échantillon.

Une fois ces ajustements visuels analysés, nous pouvons à présent, tester l'ajustement des différentes lois de probabilité citées précédemment, par les tests d'adéquation pris en compte en l'occurrence : khi-deux, K-S et Anderson-Darling.

Les résultats obtenus sont présentés dans le tableau suivant :

Tableau V.3 : Tableau résumant les résultats des tests d'adéquation pour la chronique de Bouira

Loi	Tests	Statistique du test : stat	Valeur critique : c	Décision
Normale	Khi-deux	2.625	11.02	Stat<c => ajustement accepté
	K-S	0.077	0.184	Stat<c => ajustement accepté
	A-D	0.403	2.502	Stat<c => ajustement accepté
Log-Normale	Khi-deux	7.931	11.07	Stat<c => ajustement accepté
	K-S	0.092	0.185	Stat<c => ajustement accepté
	A-D	0.294	2.502	Stat<c => ajustement accepté
Gumbel	Khi-deux	6.118	9.488	Stat<c => ajustement accepté
	K-S	0.111	0.184	Stat<c => ajustement accepté
	A-D	0.512	2.502	Stat<c => ajustement accepté
Weibull	Khi-deux	2.569	11.07	Stat<c => ajustement accepté
	K-S	0.075	0.185	Stat<c => ajustement accepté
	A-D	0.584	2.502	Stat<c => ajustement accepté
Exponentielle	Khi-deux	21.866	9.488	Stat>c => ajustement rejeté
	K-S	0.236	0.184	Stat>c => ajustement rejeté
	A-D	4.798	2.502	Stat>c => ajustement rejeté
GEV	Khi-deux	4.098	11.07	Stat<c => ajustement accepté
	K-S	0.079	0.185	Stat<c => ajustement accepté
	A-D	0.260	2.502	Stat<c => ajustement accepté
Pearson 3	Khi-deux	5.988	11.07	Stat<c => ajustement accepté
	K-S	0.088	0.185	Stat<c => ajustement accepté
	A-D	0.276	2.502	Stat<c => ajustement accepté
Log-Pearson3	Khi-deux	6.012	11.07	Stat<c => ajustement accepté
	K-S	0.085	0.185	Stat<c => ajustement accepté
	A-D	0.263	2.502	Stat<c => ajustement accepté

On remarque à travers ce tableau que les trois tests utilisés donnent des résultats positifs quant à l'ajustement de toutes les lois de probabilité étudiées.

A présent, nous allons appliquer la méthode basée sur l'harmonisation des quantiles et des temps de retour.

Les résultats obtenus pour la chronique étudiée sont les suivants :



Les quantiles calculés pour les différentes lois ainsi que les temps de retour de la plus grande valeur de la chronique étudiée sont résumés dans le tableau suivant :

**Tableau V.4 :** Tableau résumant les quantiles et les temps de retour pour la chronique de Bouira

Loi	Quantiles pour T=52 ans (mm)	Temps de retour (ans) pour P <sub>jmax</sub> =63.8 mm
Normale	56.4	444.44
Log-Normale	60.8	81.97
Gumbel	64.4	48.32
Pearson3	59.7	105.26
Log-Pearson3	59.5	100
Weibull	56.7	563.39
Exponentielle	87.2	13.30
GEV	59.5	24.88

En comparant les résultats obtenus ci-dessus, on constate que la loi Normale, Weibull et exponentielle sont à rejeter car les temps de retour et les quantiles ne correspondent pas aux valeurs observées. Par contre, les lois log-normale et Gumbel semblent être celles qui s'ajustent le mieux à notre échantillon, tandis que les lois Pearson 3 et Log-Pearson 3 donnent des temps de retour très importantes comparés à la période d'observation qui est de 52 ans.

De ce fait, on optera alors, par cette méthode, soit pour la loi de Gumbel, soit pour la loi Log-Normale.

### Conclusion:

A partir des différentes analyses effectuées, nous pouvons conclure que l'ajustement visuel comparé aux tests d'adéquation non paramétriques utilisés, permet de donner une idée plus ou moins bonne sur l'adéquation des lois de probabilité prises en compte.

Par ailleurs, il convient de signaler que l'approche d'harmonisation des quantiles et des temps de retour ainsi que l'ajustement visuel par les diagrammes quantile-quantile et le graphique par les intervalles de confiance s'avèrent ne pas être assez efficaces pour le choix de la distribution s'ajustant le mieux aux séries temporelles considérées ; en effet, le choix de la loi adéquate est basé, un tant soit peu, sur la subjectivité du modélisateur.

En définitive, nous pouvons conclure que tous ces outils classiques d'ajustement statistique des différentes lois de probabilité étudiées ne représentent pas un outil assez puissant pour la détermination de la loi s'ajustant le mieux aux chroniques considérées.

De ce fait, en vue d'affiner toutes ces approches, on aura recours à l'utilisation des méthodes bayésiennes, plus précisément les critères d'information d'Akaike et Bayésien, lesquels permettront d'opérer le choix le plus recevable entre les différentes distributions en compétition.

### V.3 Choix de la loi de probabilité adéquate par les critères d'information

#### AIC et BIC :

#### V.3.1 Résultats pour la chronique de la Station d'Ain Oussera :

Les critères AIC et BIC obtenus sont comme le montre le tableau ci-dessous :

**Tableau V.5 : Critères AIC-BIC pour la chronique d'Ain Oussera**

Loi	BIC	AIC
Log-Normale	237.583	234.781
Gumbel	237.839	235.036
Exponentielle	240.213	237.411
Pearson 3	240.640	236.436
Log-Pearson 3	240.945	236.741
GEV	241.087	236.884
Weibull	243.736	240.934
Normale	245.534	242.732

On remarque à travers ce tableau, que les deux critères sont minimaux pour la loi Log-Normale concernant les lois à deux paramètres, tandis que pour les lois à 3 paramètres, ils le sont pour la loi de Pearson 3.

Nous pouvons aussi remarquer que le critère BIC priorise la loi exponentielle par rapport à la loi Pearson 3 bien qu'elle s'ajuste mal aux valeurs extrêmes. Cela peut s'expliquer par le fait que ce critère est basé sur la parcimonie du modèle par construction. Tandis que le critère AIC classe la loi de Pearson 3 juste après les lois Log-Normale et Gumbel.

Les valeurs minimales des deux critères sont prises pour la loi Log-Normale, par conséquent, le modèle fréquentiel le plus adéquat pour la station d'Ain Oussera est la distribution Log-Normale.

### *V.3.2 Résultats pour la chronique de la Station de Bouira :*

Les critères AIC et BIC obtenus sont comme le montre le tableau ci-dessous :

**Tableau V.6:** Critères AIC- BIC pour la chronique de Bouira

Loi	BIC	AIC
Log-Normale	385.951	382.049
Gumbel	386.944	383.041
Normale	389.434	385.531
Pearson 3	389.684	383.830
Log-Pearson 3	389.779	383.925
GEV	389.881	384.027
Weibull	392.107	388.205
Exponentielle	407.619	403.716

On remarque à travers ce tableau, que les deux critères sont minimaux pour la loi Log-Normale concernant les lois à deux paramètres, tandis que pour les lois à 3 paramètres, ils le sont pour la loi de Pearson 3.

Nous pouvons aussi remarquer que le critère BIC priorise la loi normale par rapport à la loi Pearson 3 bien qu'elle soit trop sensible aux valeurs extrêmes cela est dû au fait que ce critère est basé sur la parcimonie du modèle par construction. Tandis que le critère AIC classe la loi de Pearson 3 juste après les lois Log-Normale et Gumbel.

C'est donc pratiquement la même constatation que pour la station ci-dessus étudiée à savoir celle de Bouira.

Par conséquent, vu que ces deux critères prennent la plus petite valeur pour la loi Log-Normale, on choisira alors cette dernière comme la distribution la plus adéquate à la chronique de précipitations maximales journalières de la station de Bouira.

### **V.4 Effet du changement de la taille de l'échantillon:**

Dans cette partie, nous allons étudier les conséquences engendrées par le changement du nombre d'individus composant un échantillon assez long.

Pour notre cas, on appliquera l'approche adoptée sur les séries présentées plus haut en l'occurrence, les stations d'Ain Oussera et de Bouira.

La méthode consiste à faire varier la taille de la chronique puis chercher la loi de probabilité adéquate pour chaque chronique ainsi générée qu'on comparera par la suite à la loi obtenue pour la série originale.

Ainsi, on détectera la taille à laquelle il y a changement de la loi choisie, si changement il y a. On essaiera ensuite d'en tirer les conclusions possibles.

Pour notre cas, en faisant varier la taille de l'échantillon, nous avons obtenu pour les stations précisées ci-dessus, les résultats suivants :

#### ***V.4.1 Résultats pour la chronique de la Station d'Ain Oussera :***

Nous constatons que la meilleure distribution donnée par les critères AIC et BIC est la loi Log-Normale jusqu'à l'itération ( $N=20$ ). En effet, à partir de cette taille, les critères d'information indiquent la loi exponentielle comme étant le meilleur modèle.

#### ***V.4.2 Résultats pour la chronique de la Station de Bouira :***

Pour cette station, nous avons remarqué que la meilleure distribution donnée par les critères AIC et BIC est la loi Log-Normale jusqu'à l'itération ( $N=26$ ). En effet, à partir de cette taille, la loi choisie diffère de la loi Log-Normale, car :

- ⎧ Pour  $N = \{25,26\}$ , la meilleure loi indiquée est la distribution Normale;
- ⎧ Pour  $N \in ]11,24]$ , la meilleure loi indiquée est la distribution Log – Normale; ..
- ⎧ Pour  $N = 11$ , la meilleure loi indiquée est la distribution Weibull;
- ⎧ Pour  $N \leq 9$ , la meilleure loi indiquée est la distribution exponentielle;

Par conséquent, nous ne pouvons tirer de conclusions à partir de ces résultats quant à la taille jusqu'à laquelle on peut aller lors de l'extrapolation pour les chroniques des stations étudiées.

Cela étant, par la même occasion, la conclusion à laquelle nous sommes arrivés pour les deux stations considérées, nous mène à un autre résultat intéressant qui est que pour les méthodes à seuil, celui-ci ne doit pas être choisi d'une manière arbitraire, car comme nous l'avons vu, ce n'est qu'à partir d'une certaine valeur que le modèle s'ajustant le mieux à notre chronique correspond à la loi exponentielle à deux paramètres.

De plus, l'approche selon laquelle il est possible d'estimer les quantiles pour les séries dont l'historique est court, par l'application de la méthode des excès, peut être justifiée car plus le nombre de valeurs composant la chronique diminue, plus le meilleur modèle tend vers la distribution exponentielle sachant par ailleurs, que cette dernière est un cas particulier de la distribution généralisée de Pareto utilisée pour la modélisation des séries tronquées comme mentionné dans le chapitre II du présent mémoire.

Notons que le même travail a été fait pour le reste des chroniques étudiées et nous sommes arrivées aux mêmes conclusions.

### Conclusion générale :

La notion de recherche de la loi de probabilité adéquate est une question qui a été remise à l'ordre du jour suite à la récurrence des événements extrêmes.

Les problèmes liés au dimensionnement des ouvrages, à la meilleure connaissance des zones inondables ainsi qu'au coût généré par les crues sont des demandes de nature à exiger une estimation du risque hydrologique.

Ces problématiques deviennent donc particulièrement nécessaires à résoudre afin de diminuer la vulnérabilité des zones et assurer la sécurité des populations.

L'ensemble des résultats auxquels nous sommes parvenus, à travers notre travail, nous mène aux conjectures suivantes :

En effectuant une analyse des temps de retour, nous avons pu conclure que le problème se pose essentiellement au niveau des extrêmes. Cette constatation s'est révélée vraie lors de l'analyse des ajustements visuels.

En outre, la formule la plus utilisée actuellement étant celle de Cunnane, nous avons opté pour cette formule comme formule de calcul des fréquences empiriques.

Il faut noter aussi que chacune des formules empiriques traitées à savoir : Hazen, Weibull, Chegodayev, Gringorten, Tukey et Cunnane, correspond à une certaine loi de probabilité. De ce fait, plusieurs travaux doivent être menés dans ce sens.

Pour notre cas, nous avons pu conclure que la formule de Cunnane correspond à la loi exponentielle à deux paramètres. Par la même occasion, nous pouvons retenir que l'assertion énoncée ci-dessus à savoir que la formule de Cunnane représente le meilleur compromis pour les différentes lois est confirmée.

L'ajustement graphique et les diagrammes Quantile-Quantile constituent un bon critère capable d'orienter le choix du modèle, mais ils restent insuffisants pour détecter le « quasi-vrai » modèle recherché.

Pour ce qui est des tests d'adéquation, nous pouvons émettre la même remarque que celle énoncée à l'égard des ajustements visuels, mais ils ne permettent que l'analyse de l'ajustement de la distribution considérée à l'échantillon, prise seule, sans pour autant définir la meilleure loi ; ceci d'une part. D'autre part, les théories rencontrées dans la littérature selon lesquelles le test du khi-deux ne prend en compte que les valeurs moyennes et que le test d'Anderson-Darling ne s'ajuste qu'aux valeurs extrêmes ne sont pas mises en évidence pour les séries étudiées.

La méthode se basant sur l'harmonisation des quantiles et des temps de retour met en exergue le fait qu'il faudrait désormais s'inscrire en faux contre une tradition qui veut que la plus forte valeur d'un échantillon n'ait pas une période de retour supérieure à la durée d'observation car la queue droite a une grande variabilité comme nous avons pu le constater.

Vu l'insuffisance de ces méthodes, nous avons opté pour les méthodes bayésiennes plus précisément la minimisation des critères pénalisés : AIC et BIC qui d'après la théorie mathématique selon laquelle ils ont été développés, représentent pour l'instant, l'une des meilleures façons pour choisir le modèle fréquentiel offrant le meilleur compromis entre le biais et la variance par le critère AIC et le modèle le plus parcimonieux par le critère BIC.

Ainsi, d'après ces critères, nous avons obtenu que pour les séries temporelles de précipitations maximales représentant le bassin versant de la Soummam (stations de Bouira, de Seddouk 151006 et la 150106), la meilleure distribution est la loi Log-Normale.

Le même résultat a été trouvé pour les chroniques de précipitations maximales représentant le bassin versant du Chéelif (stations d'Ain Oussera et de Tissemsilt).

Il y a donc prédominance de la loi Log-Normale pour toutes les chroniques étudiées.

Partant de là, on peut conclure qu'il faudrait dorénavant s'affranchir de l'idée selon laquelle la loi Gumbel est la meilleure distribution pour caler les précipitations maximales car ce n'est pas toujours le cas comme nous l'avons vu.

Par ailleurs, la variation de la taille de l'échantillon, nous a permis de confirmer la théorie suivant laquelle, il est suffisant de travailler sur un échantillon de 10 valeurs lors de l'application des méthodes à seuil. Cela étant, nous avons pu en déduire aussi, qu'il n'est pas possible de tirer de conclusions quant à la taille jusqu'à laquelle on pourrait aller concernant l'extrapolation.

De plus, le choix de la loi exponentielle comme modèle candidat, montre que cette dernière ne s'ajuste pas aux précipitations maximales des chroniques traitées, ce qui confirme la théorie qui fonde le principe de l'utilisation de cette distribution essentiellement pour la modélisation des séries tronquées.

En définitive, il serait donc salutaire pour les spécialistes en hydrologie statistique d'élargir leur champ d'investigation en se consacrant au développement d'autres méthodes assurant le plus possible la convergence vers le vrai modèle fréquentiel recherché en vue de résoudre cette épineuse problématique.



## Références bibliographiques

---

- [1]. Paul Meylan, Anne-Catherine Favre, André Musy (2008), Hydrologie fréquentielle : une science prédictive, Presses polytechniques et universitaires romandes ;
- [2]. Pierre Dagnelie, théorie et méthodes statistiques, volume 1, les presses agronomiques de Gembloux ;
- [3]. Bernard Bobée et Fahim Ashkar, The Gamma family and derived distributions applied in hydrology, Water Resources Publications;
- [4]. J. Bernier et R. Veron (1964), Sur quelques difficultés rencontrées dans l'estimation d'un débit de crue de probabilité donnée, Revue des statistiques appliquées, tome 12, p. 25-48 ;
- [5]. Thierry Verdel (2007) Cours : Décision et prévision statistiques, Ecole des Mines de Nancy ;
- [6]. Séries chronologiques Volume 1 : cours et exercices Université Paris-Sud, Ingénierie Mathématiques ;
- [7]. J. JACQUET (1964) Les méthodes statistiques dans les calculs de prédétermination des crues, Revue des statistiques appliquées, tome 12, p. 49-61 ;
- [8]. Klara Persson and Jesson Rydén, Exponentiated Gumbel Distribution for Estimation of Return Levels of Significant Wave Height, Journal of Environmental Statistics ;
- [9]. Martine Baillon, Mise en œuvre de lois statistiques dans la production opérationnelle de durées de retour de pluies intenses à METEO-France, DCLim/HYDRO ;
- [10]. Picco Rakotomalala, Probabilités et Statistique, Notes de cours, Université Lumière Lyon 2 ;
- [11]. Emilie Lebarbier et Tristan Mary-Huard, Septembre 2004, Le critère BIC : fondements théoriques et interprétation, Rapport de recherche, INRIA ;
- [12]. Amor BELHEDI, FSHS, Les chroniques, les distributions temporelles, Université de Tunis ;
- [13]. Pierre Ailliot, 14 mars 2011, Notes de cours sur les valeurs extrêmes ;

- [14]. D. Mouchiroud (2003), Tests d'hypothèse, Mathématiques-Deug SV-UCBL ;
- [15]. Yahiaoui, Inondations torrentielles, Cartographie des Zones Vulnérables, Chapitre 3 : Analyse fréquentielle des débits instantanés ;
- [16]. Caissie, D, Hydrology of the petitcodiac River basin in new Brunswick, Department of Fisherues and Oceans Science Branch ;
- [17]. Frédéric Bertrand et Myriam Bertrand (2012), Choix du modèle, IRIMA Université de Strasbourg, France, Econométrie Appliquée ;
- [18]. Christophe Hurlin, Séries temporelles, UFR Economie Appliquée ;
- [19]. Michel Lubrano, Septembre 2008, Test de Racine Unitaire ;
- [20]. (2007), Durée de retour de précipitations extrêmes, Météo France, Direction de la Climatologie, Annexe, Version 1.2 ;
- [21]. Pascal Massart (2010), Sélection de modèle : de la théorie à la pratique, Université de Paris-Sud, Orsay ;
- [22]. V. Klemes (1975), Applications de l'hydrologie à la gestion des ressources en eau (Au niveau de la planification et de la conception), Hydrologie opérationnelle N4, Organisation Météorologique Mondiale ;
- [23]. Aurélie Muller (2007), Comportement asymptotique de la distribution des pluies extrêmes en France, Thèse de Doctorat, Université Montpellier II.

Annexe I-1- Graphiques d'ajustement visuel pour les chroniques annexées

Annexe I-1-1 Graphiques d'ajustement visuel pour la chronique de la station Tissemsilt

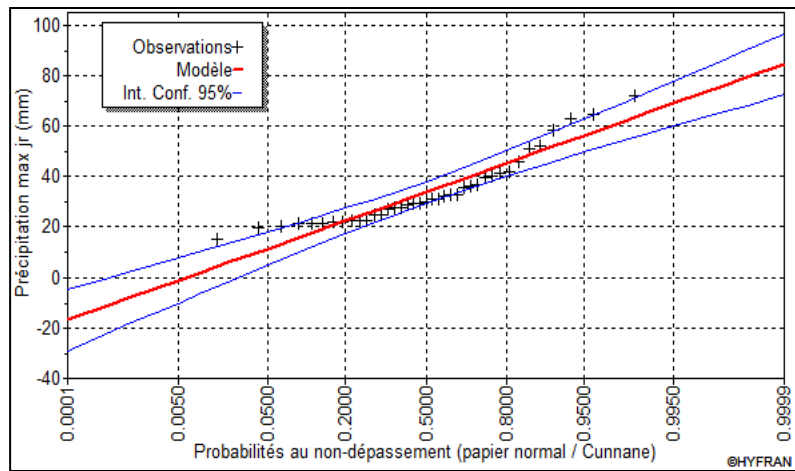


Figure I-1-1-1: ajustement graphique de la loi Normale

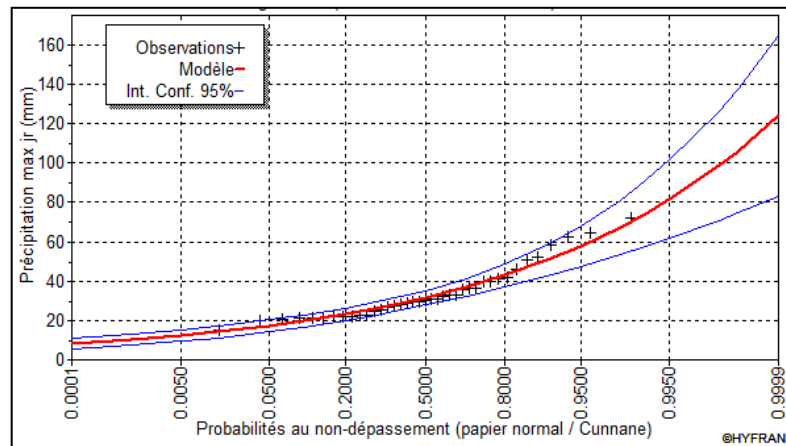


Figure I-1-1-2: ajustement graphique de la loi Log-Normale

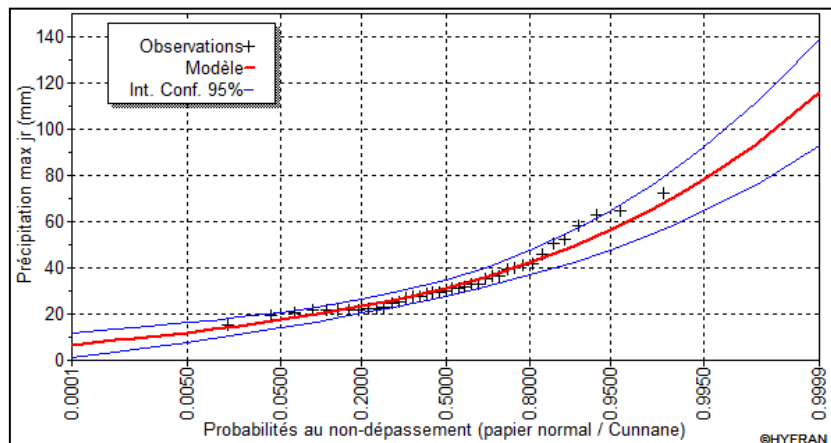


Figure I-1-1-3: Ajustement graphique de la loi Gumbel

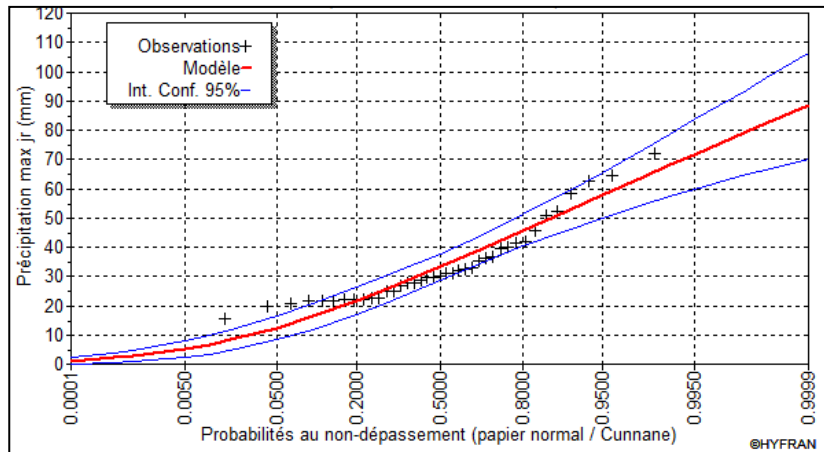


Figure I-1-1-4: ajustement graphique de la loi Weibull

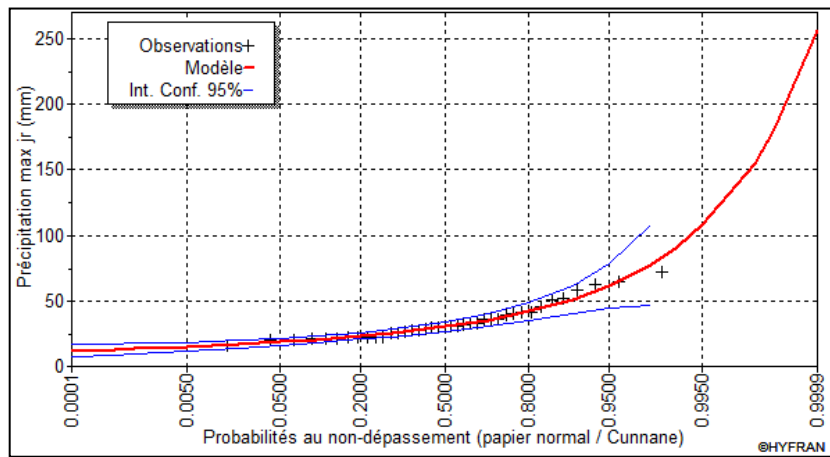


Figure I-1-1-5: ajustement graphique de la loi GEV

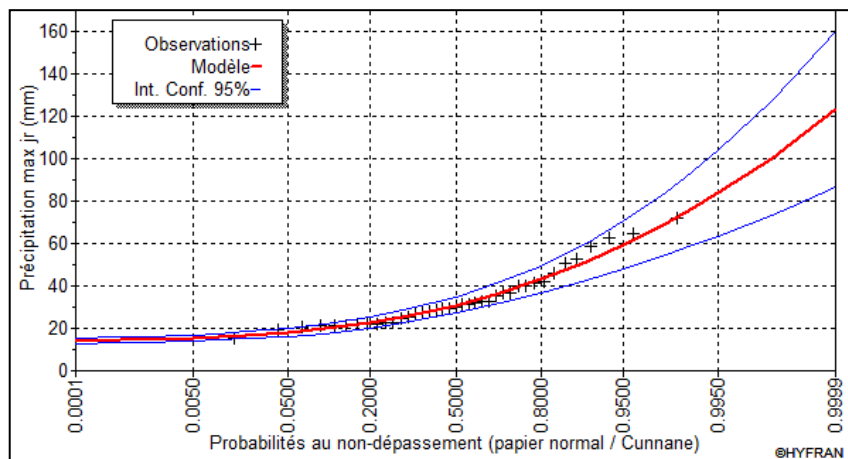


Figure I-1-1-6: ajustement graphique de la loi Pearson 3

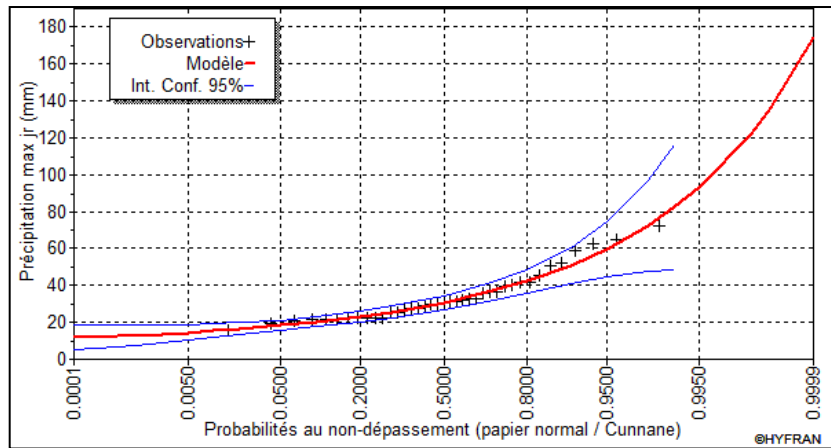


Figure I-1-1-7: ajustement graphique de la loi Log-Pearson 3

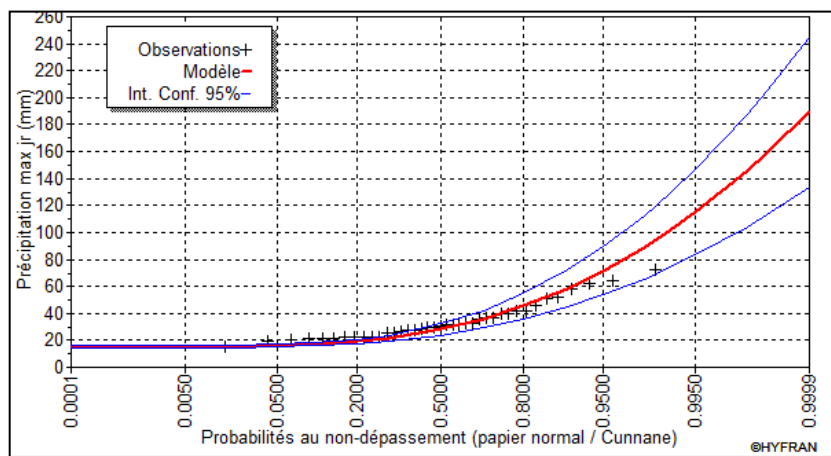


Figure I-1-1-8: ajustement graphique de la loi exponentielle

Annexe I-1-2 Graphiques d'ajustement visuel pour la chronique de la station de 150106

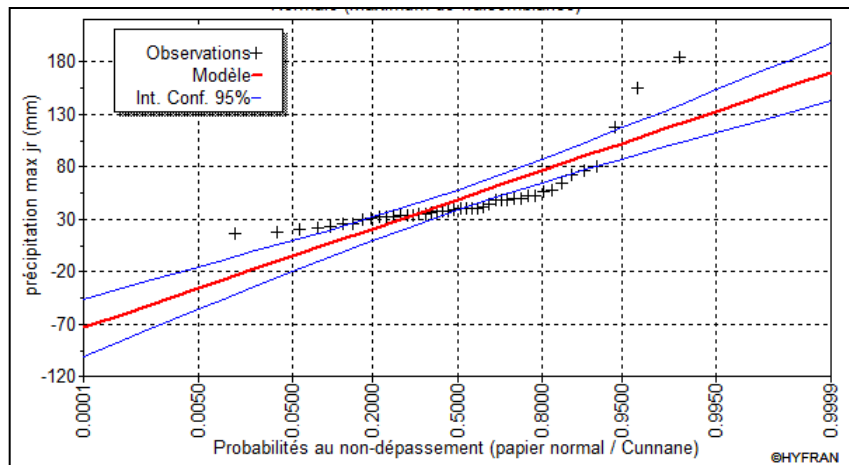


Figure I-1-2-1: ajustement graphique de la loi Normale

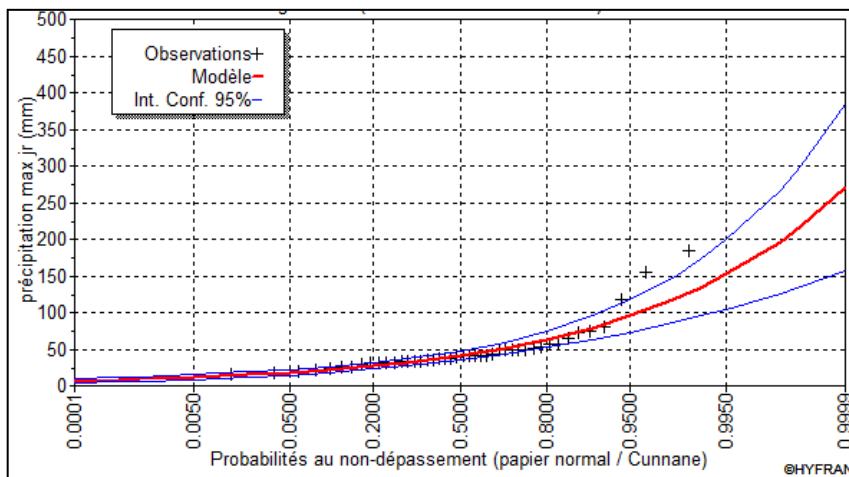


Figure I-1-2-2: ajustement graphique de la loi Log-Normale

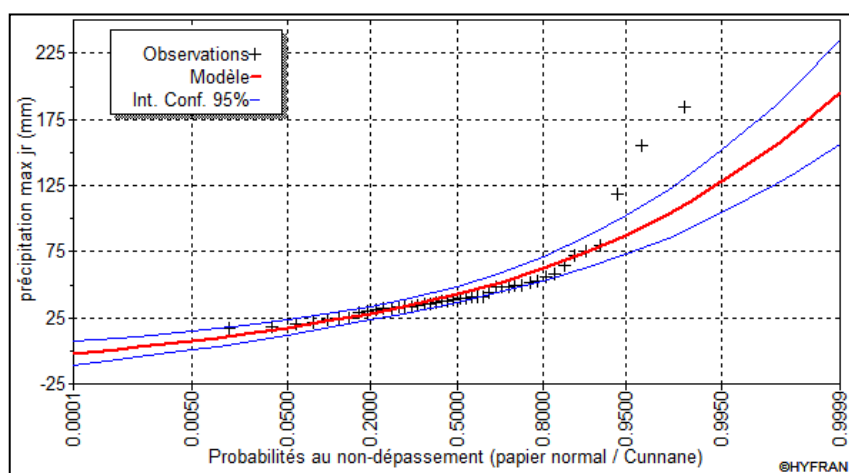


Figure I-1-2-3: ajustement graphique de la loi Gumbel

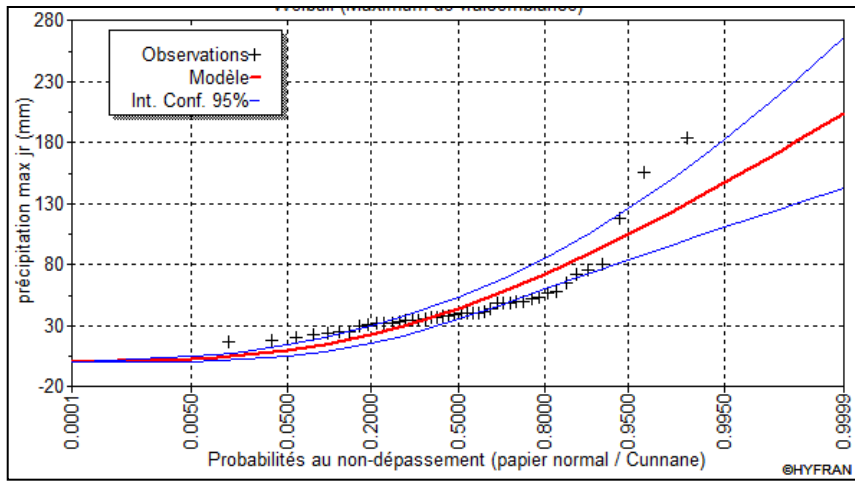


Figure I-1-2-4: ajustement graphique de la loi Weibull

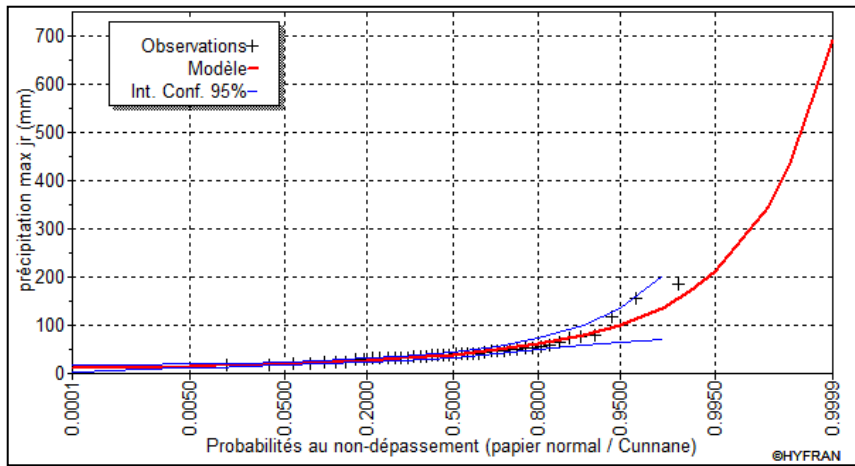


Figure I-1-2-5: ajustement graphique de la loi GEV

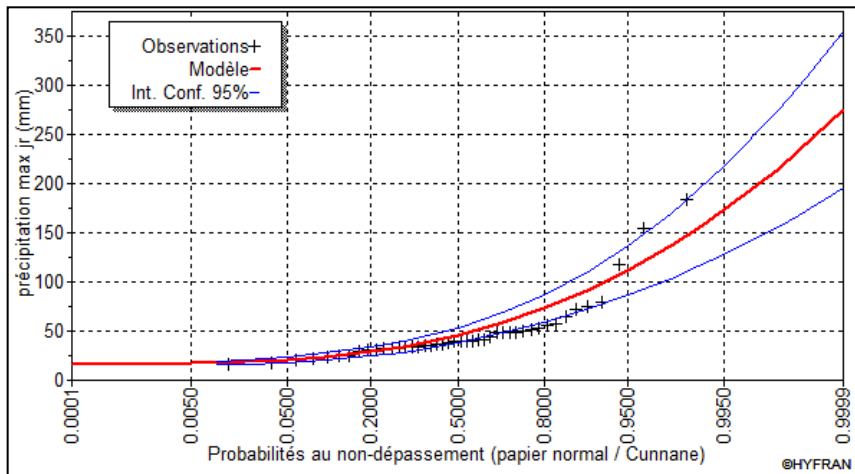


Figure I-1-2-6: ajustement graphique de la loi Pearson 3

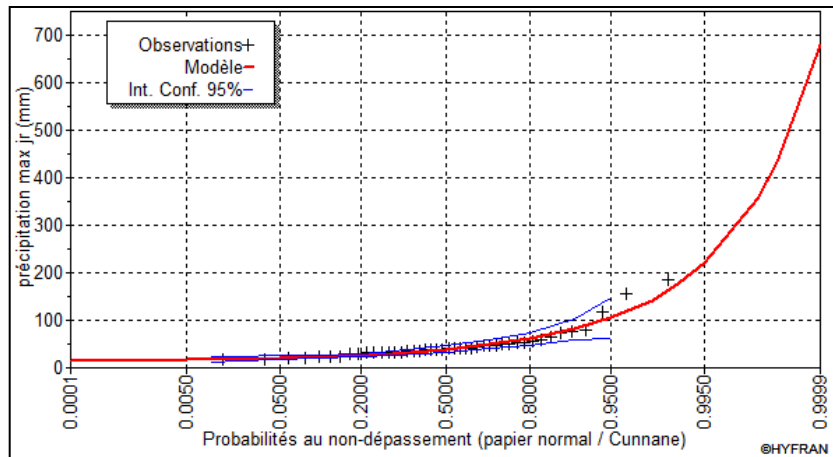


Figure I-1-2-7: ajustement graphique de la loi Log-Pearson 3

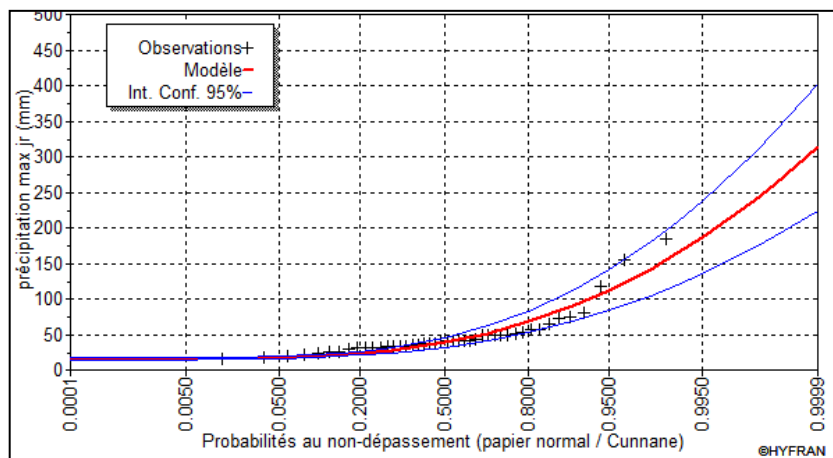


Figure I-1-2-8: ajustement graphique de la loi exponentielle



Annexe I-1-3 Graphiques d'ajustement visuel pour la chronique de la station  
Seddouk 151006

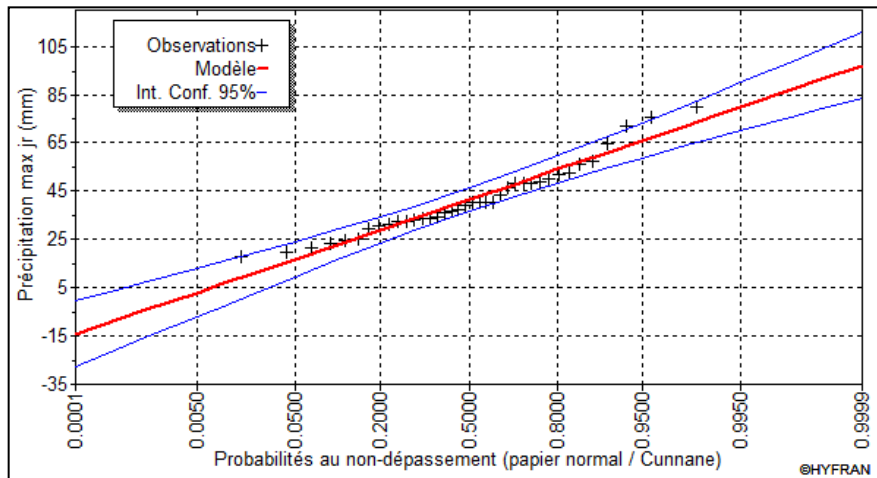


Figure I-1-3-1: ajustement graphique de la loi Normale

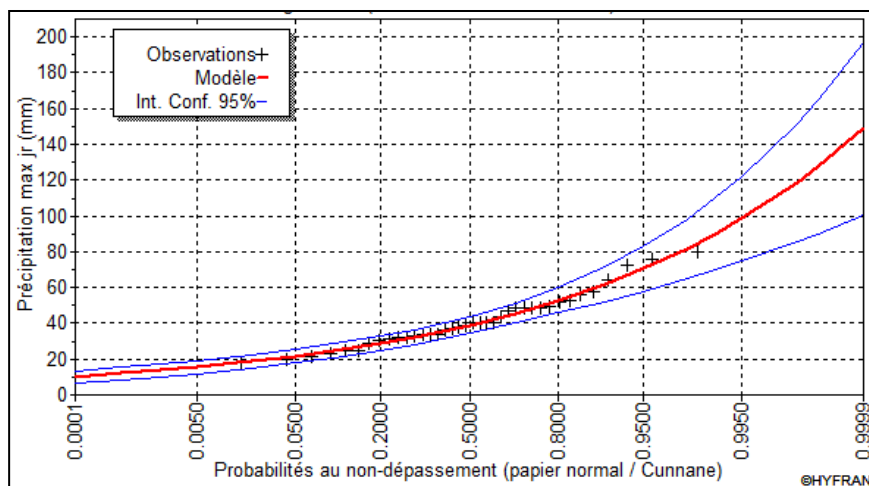


Figure I-1-3-2: ajustement graphique de la loi Log-normale

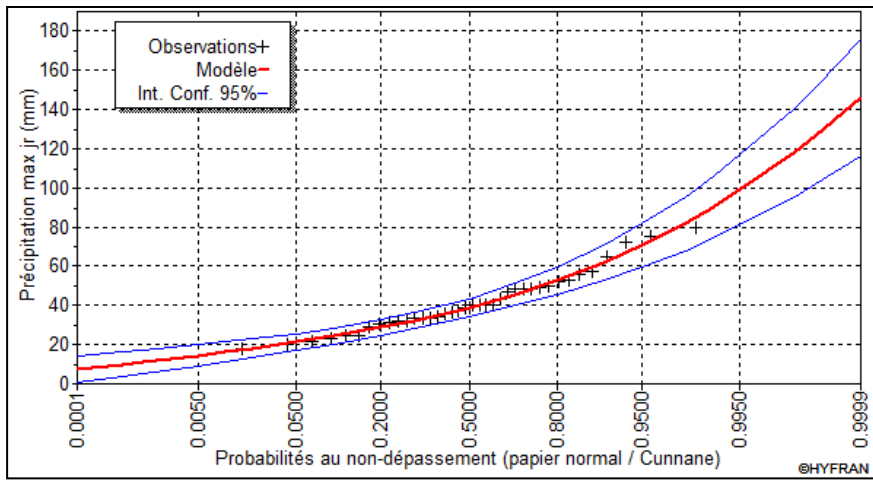


Figure I-1-3-3: ajustement graphique de la loi Gumbel

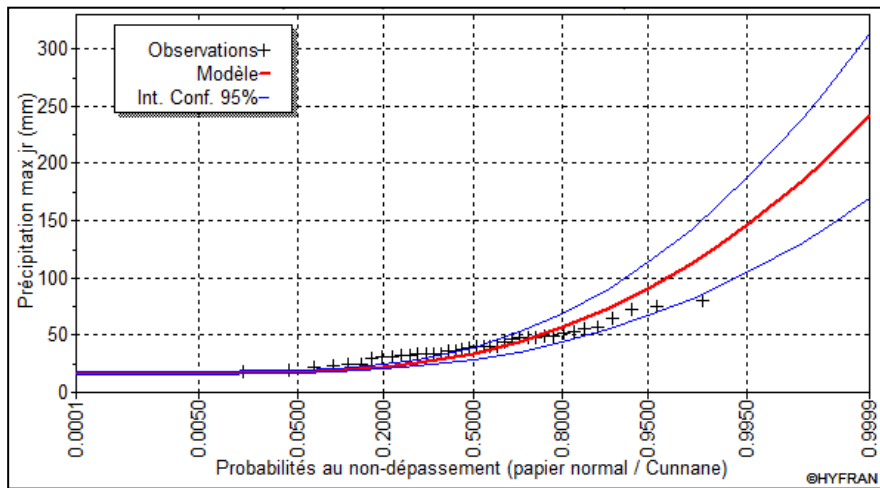


Figure I-1-3-4: ajustement graphique de la loi Weibull

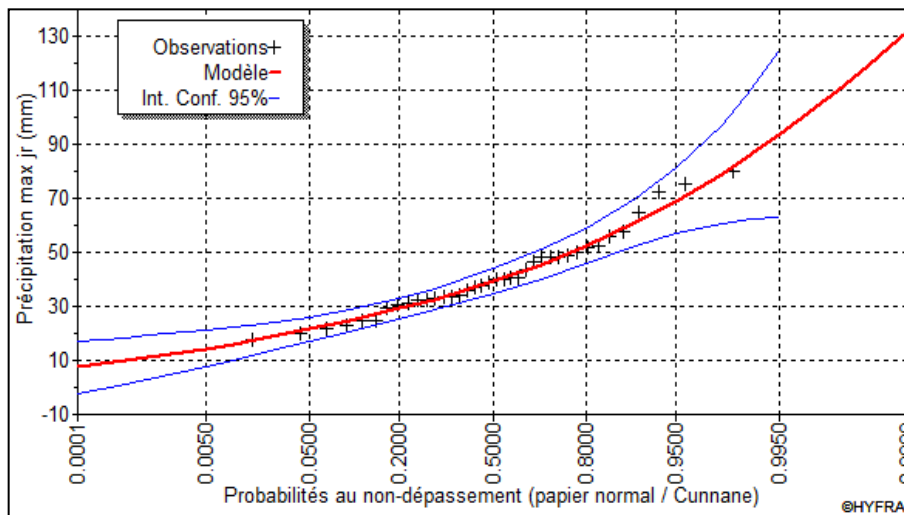


Figure I-1-3-5: ajustement graphique de la loi GEV

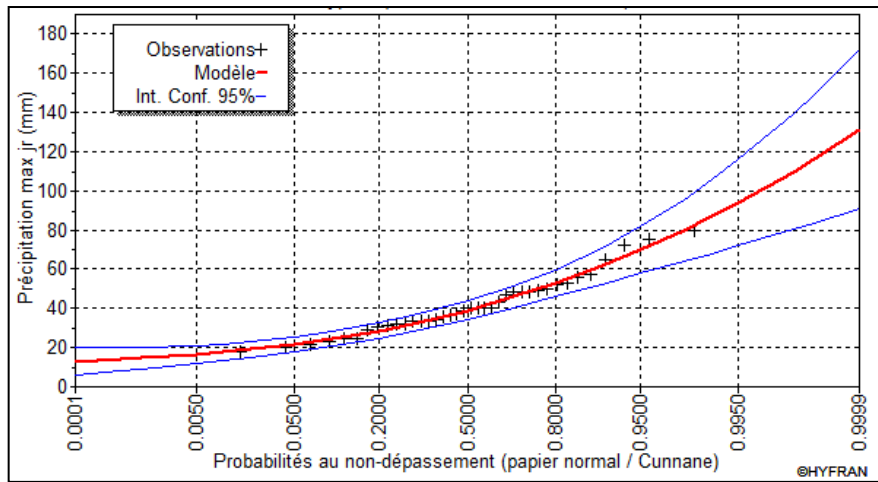


Figure I-1-3-6: ajustement graphique de la loi Pearson 3

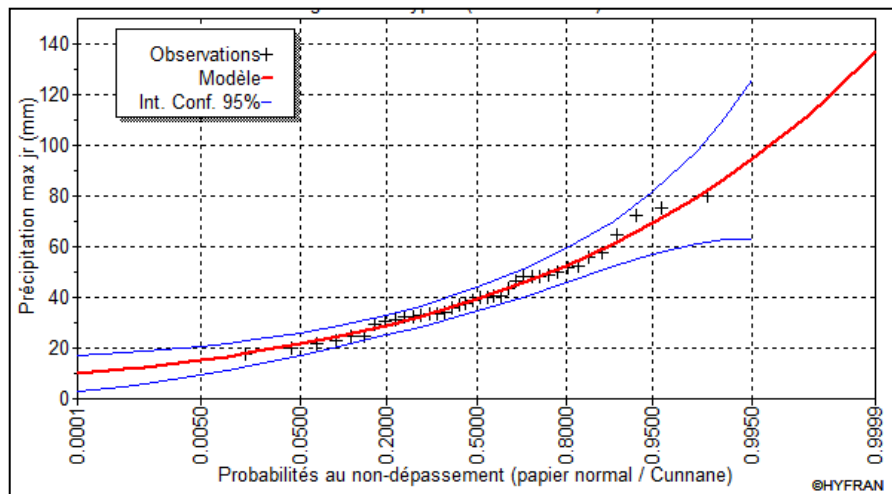


Figure I-1-3-7: ajustement graphique de la loi Log-Pearson 3

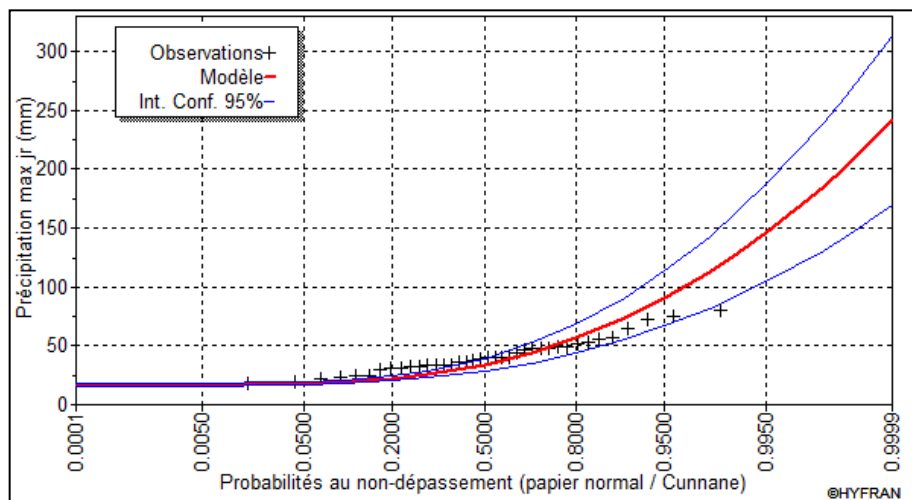


Figure I-1-3-8: ajustement graphique de la loi exponentielle

Annexe I-2 : Q-Q plot pour les différentes chroniques annexées

Annexe I-2-1 : Q-Q plot pour la chronique de la station Tissemsilt

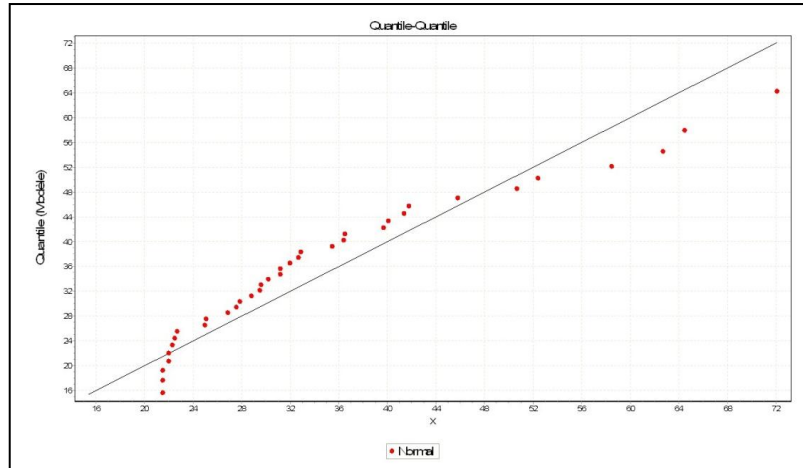


Figure I-2-1-1: Q-Q plot pour la loi Normale

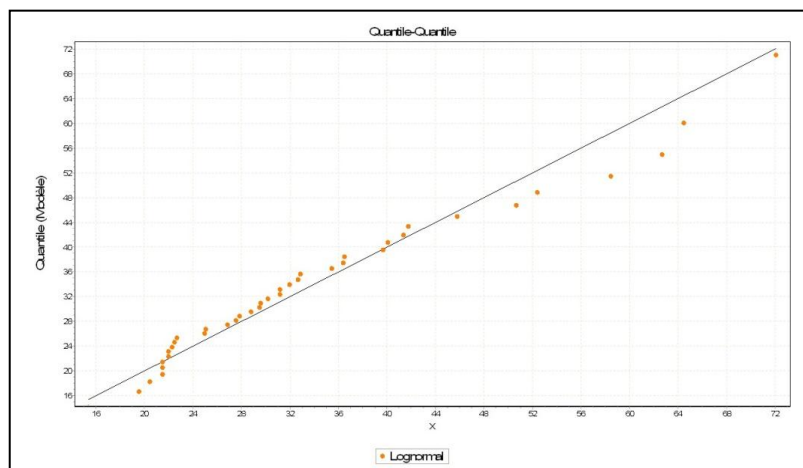


Figure I-2-1-2: Q-Q plot pour la loi Log-Normale

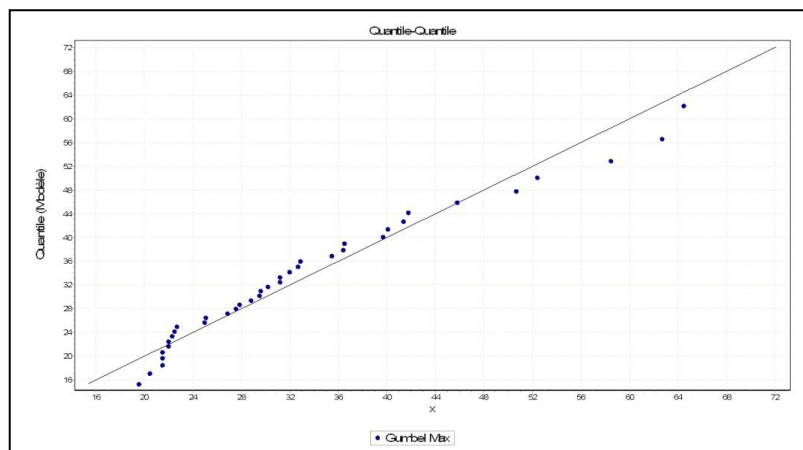


Figure I-2-1-3: Q-Q plot pour la loi Gumbel

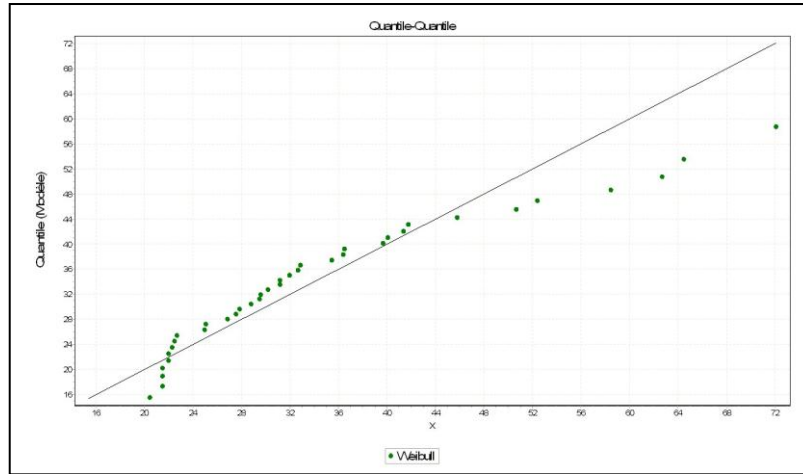


Figure I-2-1-4: Q-Q plot pour la loi Weibull

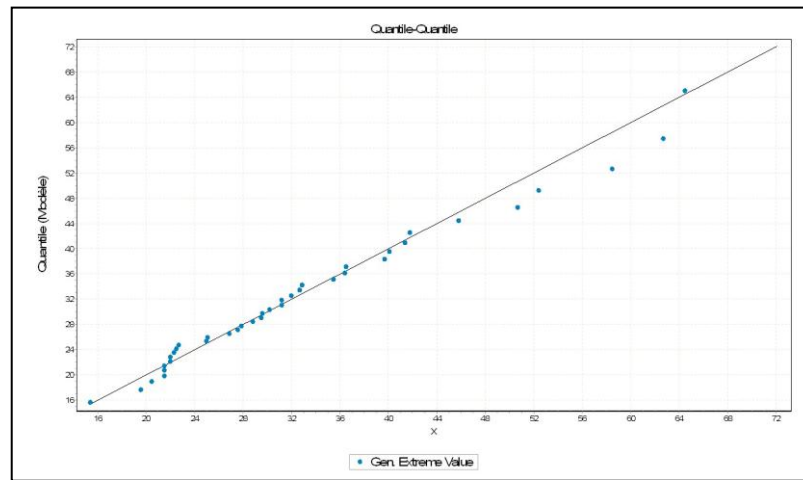


Figure I-2-1-5: Q-Q plot pour la loi GEV

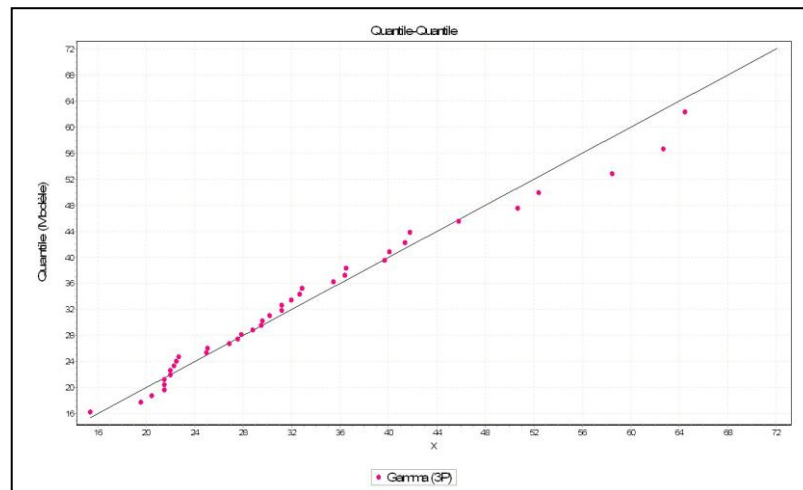


Figure I-2-1-6: Q-Q plot pour la loi Pearson 3

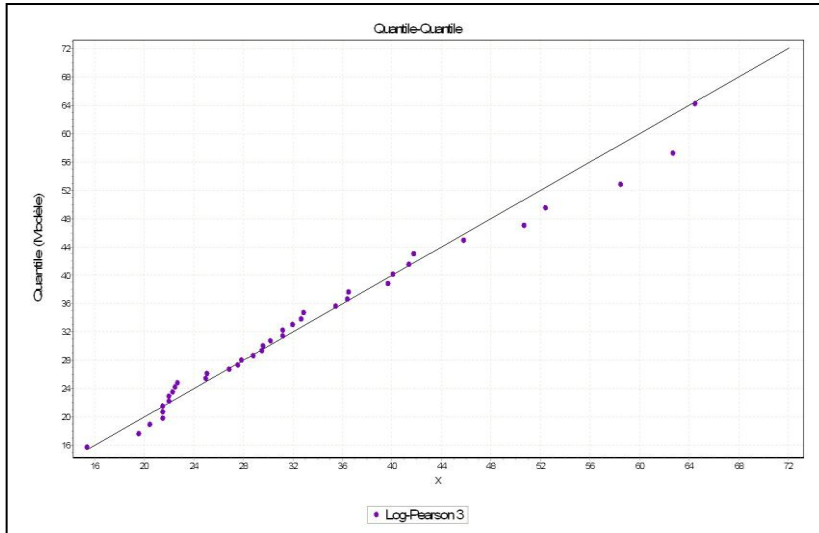


Figure I-2-1-7: Q-Q plot pour la loi Log-Pearson 3

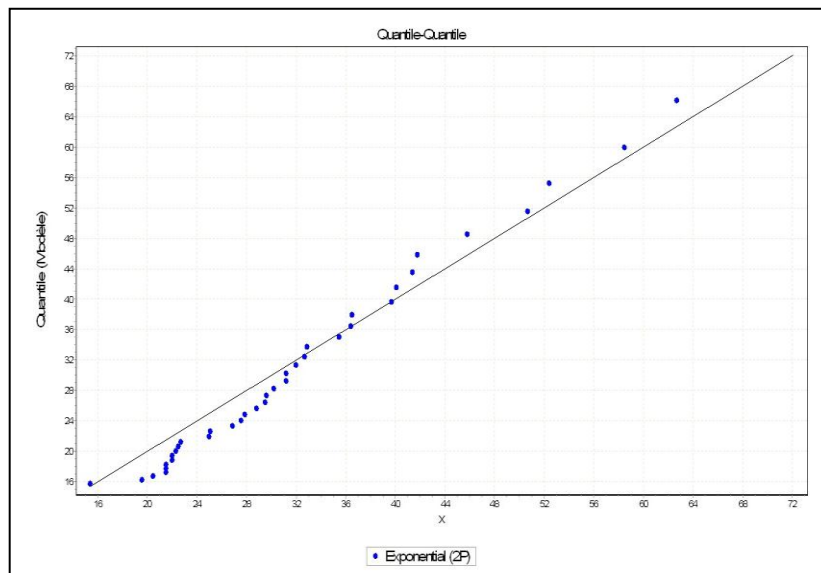


Figure I-2-1-8: Q-Q plot pour la loi exponentielle

Annexe I-2-2 : Q-Q plot pour la chronique de la station Seddouk 151006

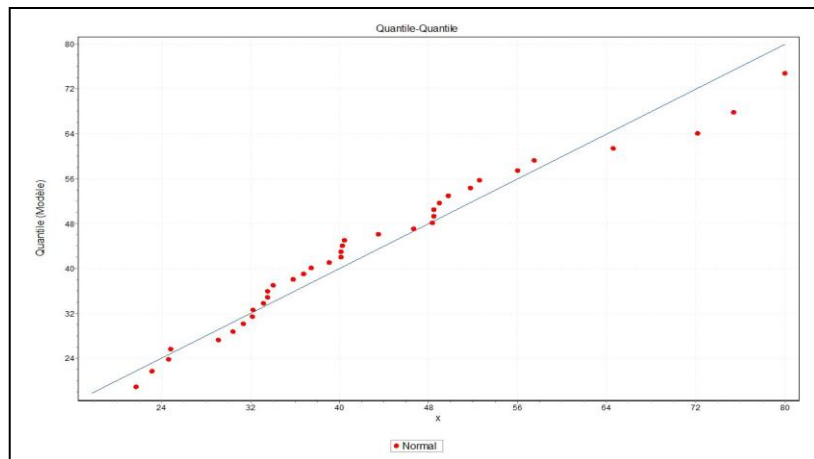


Figure I-2-2-1: Q-Q plot pour la loi Normale

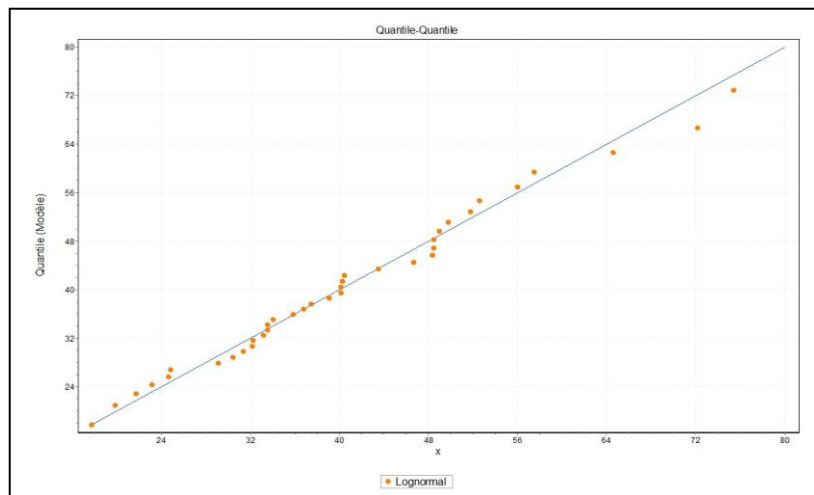


Figure I-2-2-2: Q-Q plot pour la loi Log-Normale

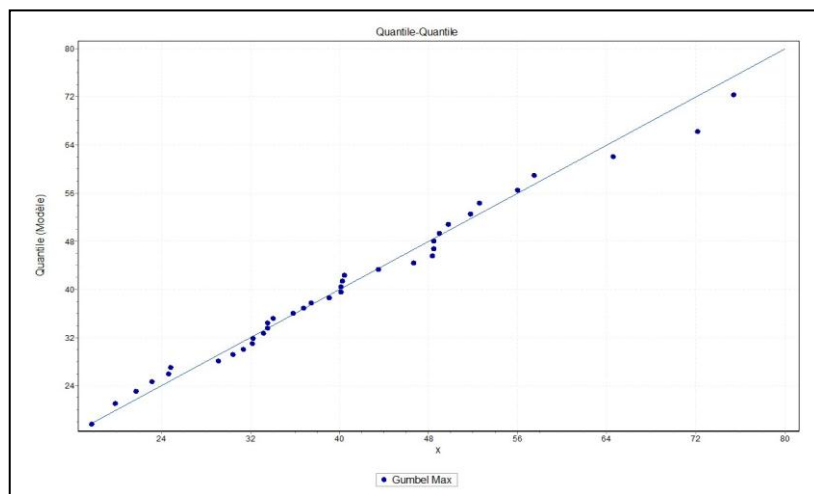


Figure I-2-2-3: Q-Q plot pour la loi Gumbel

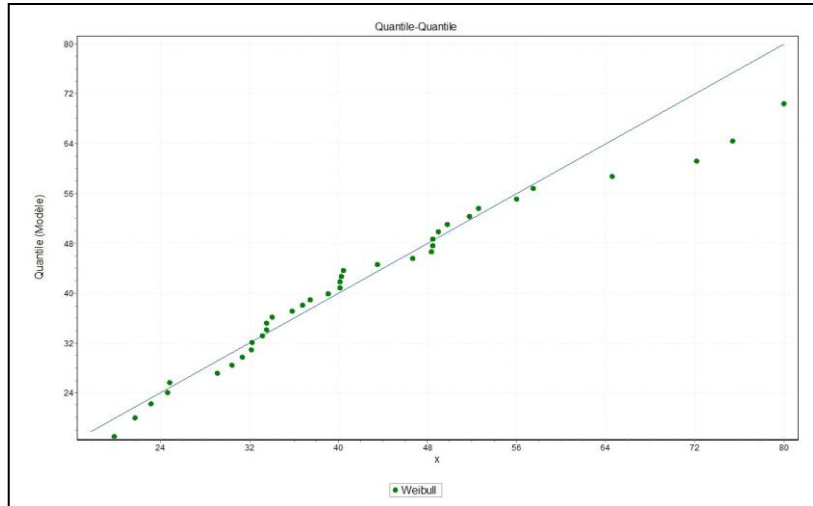


Figure I-2-2-4: Q-Q plot pour la loi Weibull

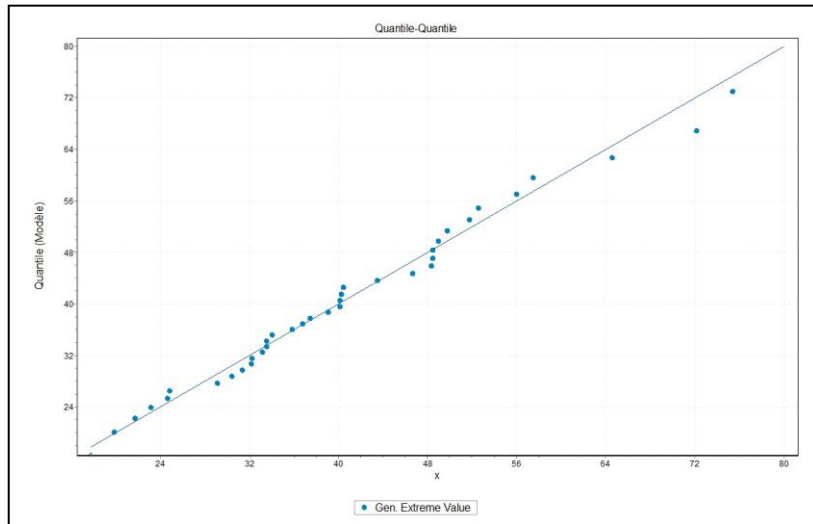


Figure I-2-2-5: Q-Q plot pour la loi GEV

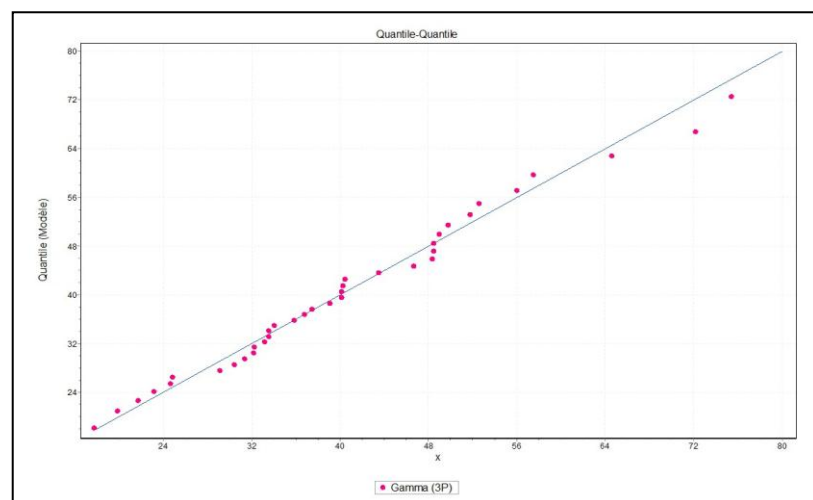


Figure I-2-2-6: Q-Q plot pour la loi Pearson 3



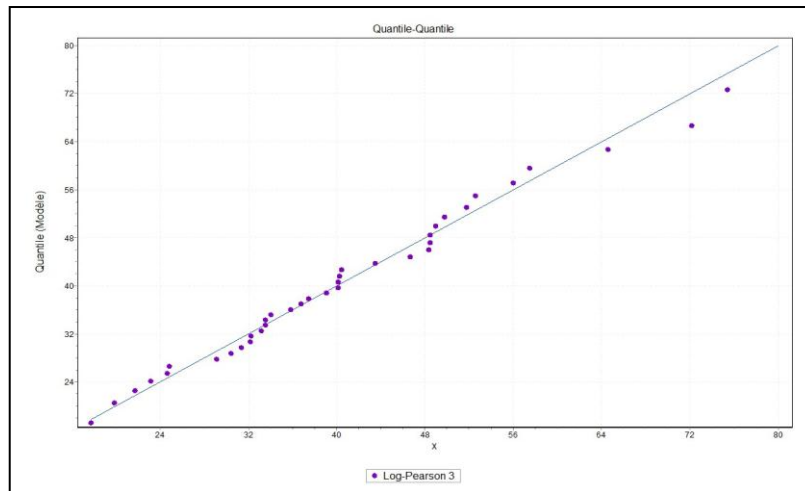


Figure I-2-2-7: Q-Q plot pour la loi Log-Pearson 3

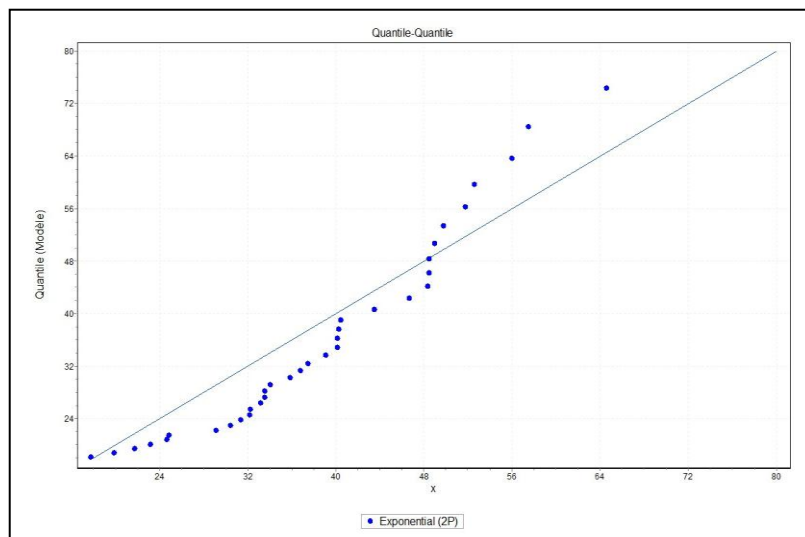


Figure I-2-2-8: Q-Q plot pour la loi exponentielle

Annexe I-2-3 : Q-Q plot pour la chronique de la station 150106

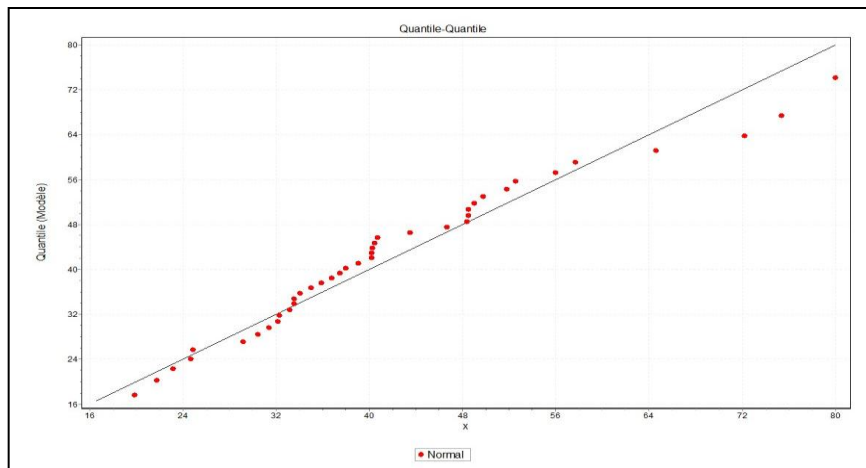


Figure I-2-3-1: Q-Q plot pour la loi Normale

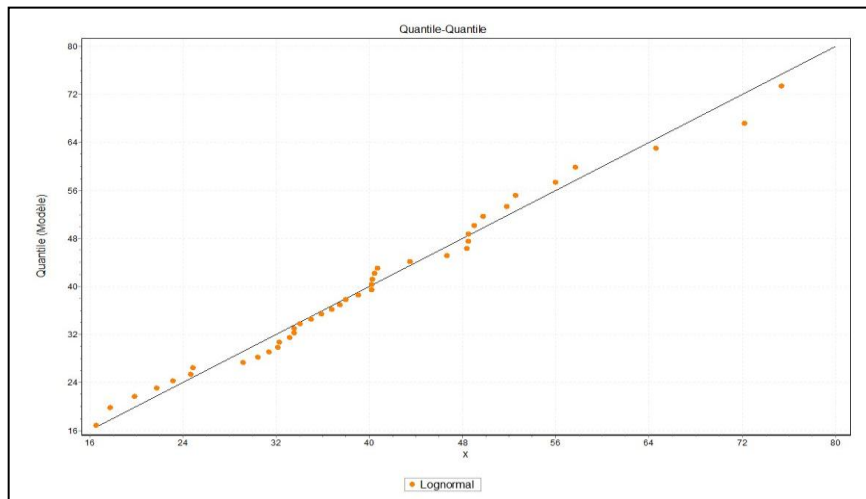


Figure I-2-3-2: Q-Q plot pour la loi Log-Normale

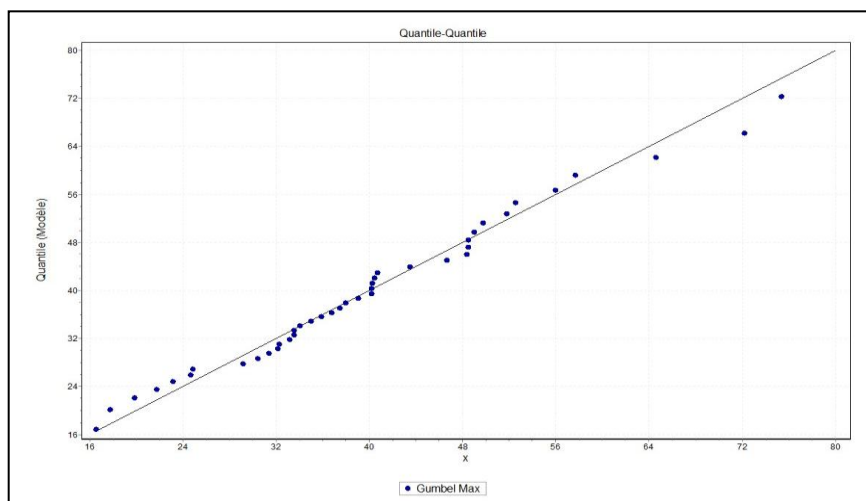


Figure I-2-3-3: Q-Q plot pour la loi Gumbel

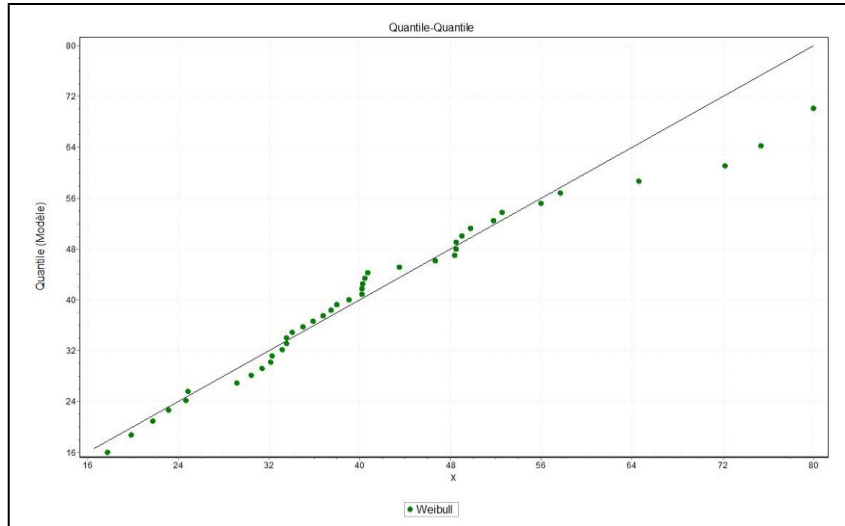


Figure I-2-3-4: Q-Q plot pour la loi Weibull

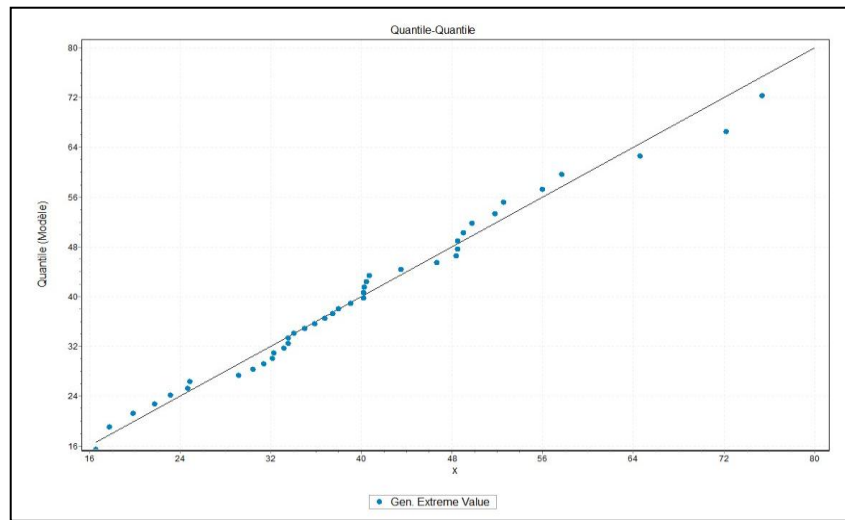


Figure I-2-3-5: Q-Q plot pour la loi GEV

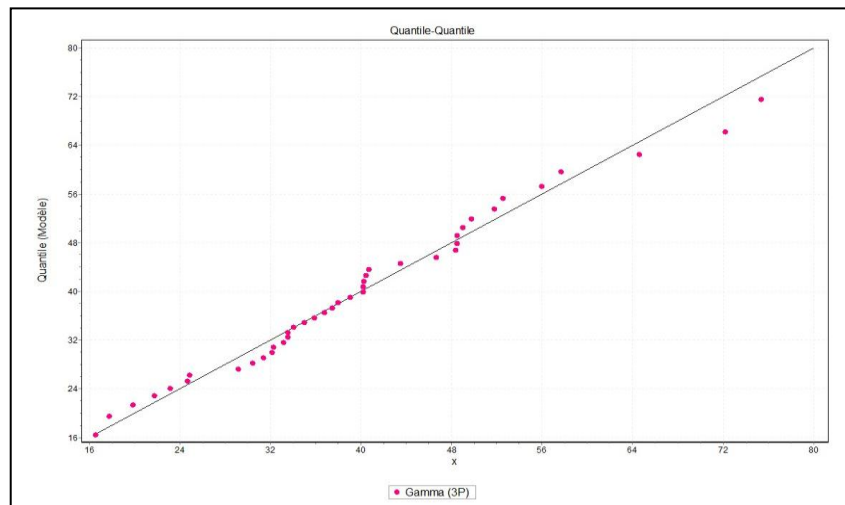
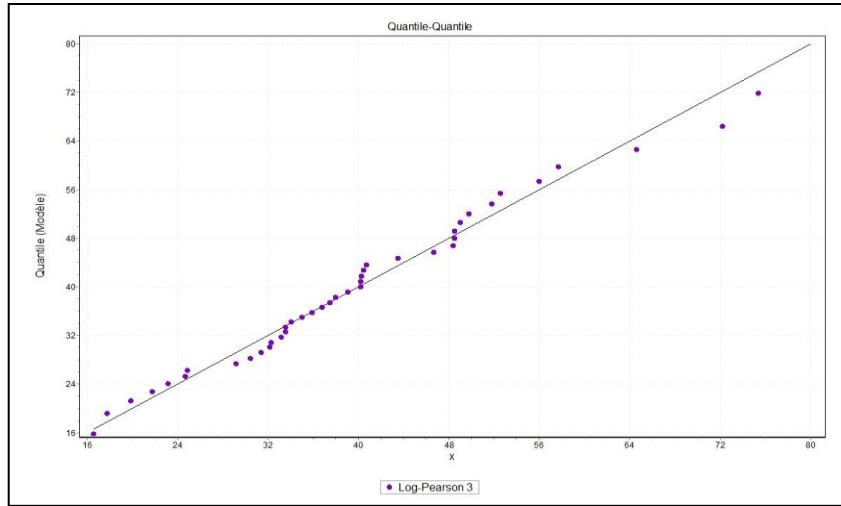
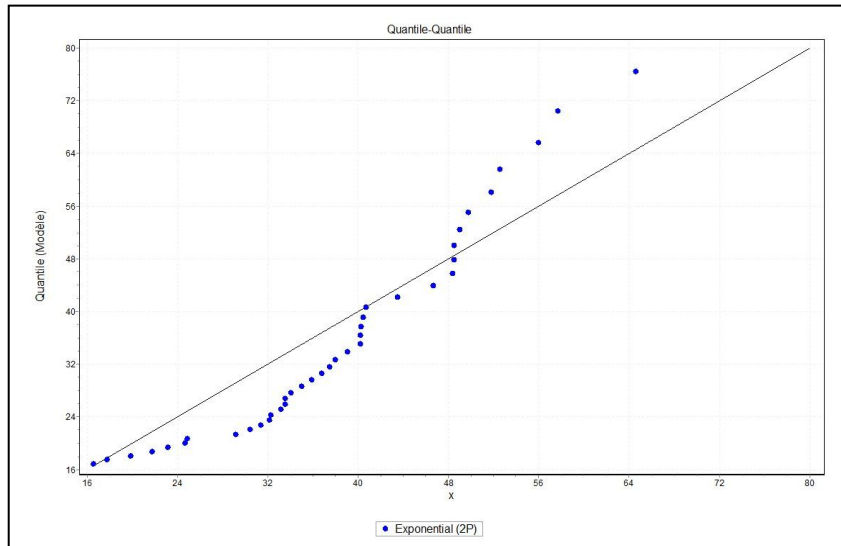


Figure I-2-3-6: Q-Q plot pour la loi Pearson 3



**Figure I-2-3-7: Q-Q plot pour la loi Log-Pearson 3**



**Figure I-2-3-8: Q-Q plot pour la loi exponentielle**

Annexe II : Résultats des tests d'adéquation pour les différentes chroniques annexées

Annexe II-1 : Résultats des tests d'adéquation pour la chronique de la station Tissemsilt

Loi	Tests	Statistique du test : stat	Valeur critique: c	Décision
Normale	Khi-deux	<b>1.893</b>	9.488	<b>Stat&lt;c =&gt; ajustement accepté</b>
	K-S	<b>0.169</b>	0.213	<b>Stat&lt;c =&gt; ajustement accepté</b>
	A-D	<b>1.546</b>	2.502	<b>Stat&lt;c =&gt; ajustement accepté</b>
Log-Normale	Khi-deux	<b>2.504</b>	7.815	<b>Stat&lt;c =&gt; ajustement accepté</b>
	K-S	<b>0.100</b>	0.213	<b>Stat&lt;c =&gt; ajustement accepté</b>
	A-D	<b>0.516</b>	2.502	<b>Stat&lt;c =&gt; ajustement accepté</b>
Gumbel	Khi-deux	<b>1.186</b>	7.815	<b>Stat&lt;c =&gt; ajustement accepté</b>
	K-S	<b>0.099</b>	0.213	<b>Stat&lt;c =&gt; ajustement accepté</b>
	A-D	<b>0.532</b>	2.502	<b>Stat&lt;c =&gt; ajustement accepté</b>
Weibull	Khi-deux	<b>3.638</b>	9.488	<b>Stat&lt;c =&gt; ajustement accepté</b>
	K-S	<b>0.134</b>	0.213	<b>Stat&lt;c =&gt; ajustement accepté</b>
	A-D	<b>1.567</b>	2.502	<b>Stat&lt;c =&gt; ajustement accepté</b>
Expo	Khi-deux	<b>5.111</b>	9.488	<b>Stat&lt;c =&gt; ajustement accepté</b>
	K-S	<b>0.204</b>	0.213	<b>Stat&lt;c =&gt; ajustement accepté</b>
	A-D	<b>2.411</b>	2.502	<b>Stat&lt;c =&gt; ajustement accepté</b>
GEV	Khi-deux	<b>0.618</b>	11.07	<b>Stat&lt;c =&gt; ajustement accepté</b>
	K-S	<b>0.093</b>	0.213	<b>Stat&lt;c =&gt; ajustement accepté</b>
	A-D	<b>0.284</b>	2.502	<b>Stat&lt;c =&gt; ajustement accepté</b>
Pearson 3	Khi-deux	<b>0.211</b>	11.07	<b>Stat&lt;c =&gt; ajustement accepté</b>
	K-S	<b>0.086</b>	0.213	<b>Stat&lt;c =&gt; ajustement accepté</b>
	A-D	<b>0.352</b>	2.502	<b>Stat&lt;c =&gt; ajustement accepté</b>
Log-Pearson3	Khi-deux	<b>1.014</b>	9.488	<b>Stat&lt;c =&gt; ajustement accepté</b>
	K-S	<b>0.095</b>	0.213	<b>Stat&lt;c =&gt; ajustement accepté</b>
	A-D	<b>0.298</b>	2.502	<b>Stat&lt;c =&gt; ajustement accepté</b>

Tableau II-1 : Résultats des tests d'adéquation pour la chronique de la station Tissemsilt

Annexe II-2: Résultats des tests d'adéquation pour la chronique de la station  
Seddouk 151006

Loi	Tests	Statistique du test : stat	Valeur critique: c	Décision
Normale	Khi-deux	2,417	11,07	<b>Stat&lt;c =&gt; ajustement accepté</b>
	K-S	0,132	0,215	<b>Stat&lt;c =&gt; ajustement accepté</b>
	A-D	0,599	2.502	<b>Stat&lt;c =&gt; ajustement accepté</b>
Log-Normale	Khi-deux	0,773	11,07	<b>Stat&lt;c =&gt; ajustement accepté</b>
	K-S	0,070	0,215	<b>Stat&lt;c =&gt; ajustement accepté</b>
	A-D	0,196	2.502	<b>Stat&lt;c =&gt; ajustement accepté</b>
Gumbel	Khi-deux	1,746	9,488	<b>Stat&lt;c =&gt; ajustement accepté</b>
	K-S	0,075	0,215	<b>Stat&lt;c =&gt; ajustement accepté</b>
	A-D	0,203	2.502	<b>Stat&lt;c =&gt; ajustement accepté</b>
Weibull	Khi-deux	3,282	9,488	<b>Stat&lt;c =&gt; ajustement accepté</b>
	K-S	0,101	0,215	<b>Stat&lt;c =&gt; ajustement accepté</b>
	A-D	0,596	2.502	<b>Stat&lt;c =&gt; ajustement accepté</b>
Expo	Khi-deux	11,798	9,488	<b>Stat&gt;c =&gt; ajustement rejeté</b>
	K-S	0,230	0,215	<b>Stat&gt;c =&gt; ajustement rejeté</b>
	A-D	3,065	2.502	<b>Stat&gt;c =&gt; ajustement rejeté</b>
GEV	Khi-deux	1,824	11.07	<b>Stat&lt;c =&gt; ajustement accepté</b>
	K-S	0,067	0,215	<b>Stat&lt;c =&gt; ajustement accepté</b>
	A-D	0,184	2.502	<b>Stat&lt;c =&gt; ajustement accepté</b>
Pearson 3	Khi-deux	0,735	11,07	<b>Stat&lt;c =&gt; ajustement accepté</b>
	K-S	0,066	0,215	<b>Stat&lt;c =&gt; ajustement accepté</b>
	A-D	0,205	2.502	<b>Stat&lt;c =&gt; ajustement accepté</b>
Log-Pearson3	Khi-deux	1,772	11,07	<b>Stat&lt;c =&gt; ajustement accepté</b>
	K-S	0,069	0,215	<b>Stat&lt;c =&gt; ajustement accepté</b>
	A-D	0,188	2.502	<b>Stat&lt;c =&gt; ajustement accepté</b>

Tableau II-2: Résultats des tests d'adéquation pour la chronique de la station Seddouk 151006

Annexe II-3: Résultats des tests d'adéquation pour la chronique de la station 150106

Loi	Tests	Statistique du test : stat	Valeur critique: c	Décision
Normale	Khi-deux	5,270	9,488	Stat>c => ajustement rejeté
	K-S	0,141	0,205	Stat>c => ajustement rejeté
	A-D	0,647	2,502	Stat<c => ajustement accepté
Log-Normale	Khi-deux	1,485	11,07	Stat<c => ajustement accepté
	K-S	0,081	0,205	Stat<c => ajustement accepté
	A-D	0,287	2,502	Stat<c => ajustement accepté
Gumbel	Khi-deux	1,739	11,07	Stat<c => ajustement accepté
	K-S	0,0719	0,205	Stat<c => ajustement accepté
	A-D	0,276	2,502	Stat<c => ajustement accepté
Weibull	Khi-deux	1,499	9,488	Stat<c => ajustement accepté
	K-S	0,110	0,205	Stat<c => ajustement accepté
	A-D	0,605	2,502	Stat<c => ajustement accepté
Expo	Khi-deux	10,245	7,815	Stat>c => ajustement rejeté
	K-S	0,248	0,205	Stat>c => ajustement rejeté
	A-D	3,593	2,502	Stat>c => ajustement rejeté
GEV	Khi-deux	1,768	11,07	Stat<c => ajustement accepté
	K-S	0,080	0,205	Stat<c => ajustement accepté
	A-D	0,244	2,502	Stat<c => ajustement accepté
Pearson 3	Khi-deux	1,698	11,07	Stat<c => ajustement accepté
	K-S	0,083	0,205	Stat<c => ajustement accepté
	A-D	0,269	2,502	Stat<c => ajustement accepté
Log-Pearson3	Khi-deux	1,730	11,07	Stat<c => ajustement accepté
	K-S	0,085	0,205	Stat<c => ajustement accepté
	A-D	0,256	2,502	Stat<c => ajustement accepté

Tableau II-3: Résultats des tests d'adéquation pour la chronique de la station 150106

Annexe III : Résultats de la méthode d'harmonisation des quantiles et des temps de retour des chroniques annexées

Annexe III-1 : Résultats de la méthode d'harmonisation des quantiles et des temps de retour pour la chronique de la station Tissemsilt

<b>Loi</b>	<b>Quantiles pour T=39 ans (mm)</b>	<b>Temps de retour (ans) pour Pjmax= 72.1 mm</b>
Normale	<b>60.5</b>	<b>400.00</b>
Log-Normale	<b>64.7</b>	<b>81.97</b>
Gumbel	<b>66.9</b>	<b>103.44</b>
Pearson3	<b>67.0</b>	<b>46.57</b>
Log-Pearson3	<b>69.1</b>	<b>67.59</b>
Weibull	<b>62.5</b>	<b>146.65</b>
Exponentielle	<b>84.80</b>	<b>35.36</b>
GEV	<b>73.3</b>	<b>ND</b>

Tableau III-1: quantiles et temps de retour calculés

Annexe III-2 : Résultats de la méthode d'harmonisation des quantiles et des temps de retour pour la chronique de la station Seddouk 151006

Tableau III-2: quantiles et temps de retour calculés

<b>Loi</b>	<b>Quantiles pour T=38 ans (mm)</b>	<b>Temps de retour (ans) pour Pjmax= 80mm</b>
Normale	<b>70.5</b>	<b>196.08</b>
Log-Normale	<b>78.3</b>	<b>43.86</b>
Gumbel	<b>78.5</b>	<b>42.95</b>
Pearson3	<b>76.9</b>	<b>63.47</b>
Log-Pearson3	<b>76.4</b>	<b>50.00</b>
Weibull	<b>71.8</b>	<b>165.25</b>
Exponentielle	<b>106</b>	<b>25.61</b>
GEV	<b>76.1</b>	<b>36.20</b>



Annexe III-3 : Résultats de la méthode d'harmonisation des quantiles et des temps de retour pour la chronique de la station 150106

**Tableau III-3:** quantiles et temps de retour calculés

<b>Loi</b>	<b>Quantiles pour T=41 ans (mm)</b>	<b>Temps de retour (ans) pour Pjmax= 80 mm</b>
Normale	<b>75.2</b>	<b>243.90</b>
Log-Normale	<b>89.7</b>	<b>43.86</b>
Gumbel	<b>89.5</b>	<b>46.02</b>
Pearson3	<b>84.8</b>	<b>72.03</b>
Log-Pearson3	<b>85</b>	<b>50.00</b>
Weibull	<b>76.9</b>	<b>164.76</b>
Exponentielle	<b>129</b>	<b>32.08</b>
GEV	<b>84.6</b>	<b>35.36</b>

Annexe IV : Résultats des critères AIC-BIC calculés pour les chroniques annexées

Annexe IV-1 : Résultats des critères AIC-BIC calculés pour la chronique de la station Tissemsilt

**Tableau IV-1:** critères AIC-BIC calculés

<b>Loi</b>	<b>BIC</b>	<b>AIC</b>
Log-Normale	308.341	305.014
Gumbel	309.598	306.271
Pearson 3	311.270	306.279
Log-Pearson 3	312.623	307.632
GEV	312.877	307.886
Exponentielle	314.734	311.407
Weibull	317.853	314.526
Normale	320.669	317.342

Annexe IV-2 : Résultats des critères AIC-BIC calculés pour la chronique de la station  
Seddouk 151006

**Tableau IV-2:** critères AIC-BIC calculés

Loi	BIC	AIC
Log-Normale	314.864	311.589
Gumbel	315.021	311.746
Pearson 3	318.136	313.223
Log-Pearson 3	318.426	313.513
GEV	318.571	313.658
Weibull	318.740	315.465
Normale	319.761	316.486
Exponentielle	325.912	322.637

Annexe IV-3: Résultats des critères AIC-BIC calculés pour la chronique de la station 150106

**Tableau IV-3:** critères AIC-BIC calculés

Loi	BIC	AIC
Log-Normale	339.409	335.982
Gumbel	339.422	335.995
Pearson 3	342.769	337.628
Log-Pearson 3	342.900	337.759
GEV	342.986	337.846
Weibull	343.142	339.715
Normale	344.400	340.972
Exponentielle	351.769	348.341