



REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
Université Mustapha Stambouli - Mascara
Faculté des Sciences et Technologie



Polycopié de Cours

Analyse et modélisation hydrologique

Cours, TD

**Support destiné aux étudiants de 1^{ère} année Master Ressources
Hydrauliques**

Présenté par :

Dr Amel TALIA

**Algérie
2023**

Avant Propos

Ce *polycopié* correspond au cours d'analyse et modélisation hydrologique . Il s'agit d'un polycopié basé sur les notes de cours, des présentations ainsi que d'ajouts personnels. L'objectif de ce cours est de présenter les notions fondamentales actualisées. D'une part, il permettra aux étudiants d'acquérir les connaissances nécessaires pour l'analyse rationnelle des mesures et observations portant sur les facteurs hydrométéorologiques d'un phénomène, en vue d'élucider son mécanisme et ses lois de probabilité, son objectif sera souvent de mettre sur pied des méthodes de prédétermination quantitative de l'amplitude ou de la probabilité d'occurrence du dit phénomène. D'autre part, ce document permettra d'utiliser différentes approches de modélisation hydrologique pour déterminer la relation pluie-débit, dans le but de prévision ou d'utilisation de tels modèles pour des bassins non jaugés.

Ce cours polycopié peut aussi constituer une référence utile pour les étudiants des différentes spécialités des sciences de l'ingénieur et des professionnels intéressés par la question.

Mascara, Janvier 2023.

Table des matières

Avant propos	i
Introduction.....	01
....	
Chapitre 1 : Notions de probabilité et d'analyse statistique appliquée à l'hydrologie	
1. Rappel : Terminologie et notions fondamentales	03
1.1 Vocabulaire	03
1.1.1 Les variables.....	03
1.1.2 Distribution d'une série statistique.....	04
1.1.3 La fonction de densité de probabilité	05
1.1.4 Période de retour	06
1.1.5 Notions de risque	07
1.2 Caractéristiques d'une distribution.....	08
1.2.1 Paramètres de tendance centrale.....	08
1.2.2 Paramètres de dispersion	08
1.3 Analyse fréquentielle.....	10
2. Types de fonctions de distribution en hydrologie	11
2.1 La loi normale	12
2.2 Loi log-normale.....	14
2.3 Loi de Gumbel	15
3. Valeur extrêmes d'une variable en hydrologie.....	17
4. Théorie générale de l'ajustement statistique.....	19
4.1 Estimation par la méthode des moments	19
4.2 Estimation par la méthode du maximum de vraisemblance.....	20
5. Intervalles de confiance et bandes de confiance.....	21
6. Tests d'ajustement	23
6.1 Rappels sur les principes des tests statistiques.....	23
6.2 Test chi-carré de K. Pearson.....	26
6.3 Test de Kolmogorov-Smirnov	27
Mise en application.....	28

Chapitre 2 : Corrélation et analyse des données

1. Généralités :.....	40
1.1 Association entre variable.....	40
1.2 Les données aberrantes.....	41
2. Généralités sur la corrélation	41
2.1 Notations et définitions.....	41
2.2 Analyse primaire de qualité.....	42
2.3 Complément statistique des données hydrologiques.....	43
3. Processus, variables et séries	46
3.1 Définition.....	46
3.2 Valeurs caractéristiques d'une série chronologique.....	46
3.3 Recherche de la tendance générale.....	46
3.4 Recherche diurne et composante saisonnière.....	48
3.5 Recherche de la composante cyclique.....	49
3.6 Caractère aléatoire des séries d'observations.....	49
3.6.1 Test du rapport de Von Neumann.....	49

3.6.2	Auto- corrélogramme	50
3.6.3	Test de corrélation sur le rang.....	51
3.6.4	Statistique de rang de Spearman	52
3.7	Tests de détection de ruptures	52
3.7.1	Test de Pettit « version modifiée de Mann-Withney »	52
3.7.2	Test du rapport de vraisemblance.....	53
3.7.3	Statistique U	54
3.7.4	La Procédure Bayésienne de Lee et Heghinian.....	55
3.7.5	La procédure de segmentation des séries hydrométéorologiques.....	57
3.7.6	Les conditions d'application des méthodes.....	58
3.8	Lissage par la méthode de la Moyenne Mobile.....	59
4.	Test d'homogénéité	59
4.1	Définition de la notion d'homogénéité.....	59
4.2	Approche générale.....	60
4.2.1	Tests graphiques.....	60
4.2.1.1	Le simple cumul.....	60
4.2.1.2	Test du double cumul.....	61
4.2.1.3	Méthode des résidus.....	61
4.2.2	Tests analytiques.....	62
4.2.2.1	Test de la médiane.....	62
4.2.2.2	Test de Wilcoxon ou Test des rangs.....	63
	Mise en application.....	64

Chapitre 3 : Modélisation hydrologique

1.	La modélisation hydrologique.....	76
1.1	Définition du modèle hydrologique.....	76
1.2.	Quelques éléments de vocabulaire.....	76
1.2.1.	Les variables d'un modèle hydrologique.....	76
2.	Pourquoi des modèles hydrologiques ?.....	77
3.	Différentes approches de modélisation.....	78
3.1.	Selon le degré d'abstraction Modèles physiques.....	79
3.2.	Selon la nature des variables.....	79
3.3.	Selon la discrétisation spatiale.....	80
3.4.	Selon la description des sous processus.....	80
3.5.	Choix d'un modèle.....	80
3.6.	Phase de calage et de validation d'un modèle.....	81
3.6.1.	Calage.....	81
3.6.2.	Validation	81
3.7.	Critères d'appréciation des modèles.....	81
3.8.	Etapas d'élaboration d'un modèle.....	84
4.	Présentation de quelques outils.....	84
4.1.	Le logiciel GARDÉNIA.....	84
4.2.	EMILE	84
4.3.	Le SIG.....	85
5.	Applications hydrologiques de modèles pluie-débit.....	85
5.1	Le Soil and Water Assessment Tool (SWAT).....	85
5.2.	Les modèles ' boîte noire '.....	85
	Mise en application.....	88

Sujets proposés.....	100
Conclusion.....	106
Références bibliographiques.....	107
Annexe 1 : Table de Gauss.....	113
Annexe 2 : Table de Student.....	114
Annexe 3 : Table de χ^2	115

Introduction

La définition de l'hydrologie reste difficile, car ce n'est pas une science unifiée, elle change d'une époque ou d'un hydrologue à l'autre. Etymologiquement elle est définie comme étant "science qui traite des eaux". C'est la science qui cherche à répondre à la question « Qu'arrive-t-il aux précipitations ? ». Question simple en apparence mais sur laquelle les nombreux chercheurs qui s'y sont penchés n'ont pas trouvé une réponse entièrement satisfaisante. Cela s'explique par le fait que les éléments qui font l'objet d'études en hydrologie sont eux-mêmes sujets à de nombreuses incertitudes tant dans la connaissance théorique des chercheurs que dans leurs mesures en nature. Cette complexité a donné lieu à plusieurs branches de l'hydrologie: hydrologie maritime, côtière, des bassins versants. Dans les dictionnaires, elle est parfois définie comme la science qui étudie les eaux, leurs caractéristiques, leurs propriétés, s'appuyant sur des considérations physiques, météorologiques, géologiques ou chimiques. L'objet fondamental d'étude de l'hydrologie est le cycle de l'eau. Ses thèmes d'étude sont les précipitations, l'évaporation, l'infiltration, le ruissellement, les écoulements dans les nappes et les cours d'eau, et le transport de substances dissoutes ou en suspension.

L'hydrologie est donc une science pluridisciplinaire comprenant l'hydrologie de surface, la glaciologie, l'hydrogéologie, la nivologie, la physico-chimie, en incluant aussi l'étude de l'érosion ou du transport de sédiments. Mais elle est également plus ou moins directement liée à la météorologie, l'hydraulique, la géographie, la géologie, la biologie ou l'écologie. Il est donc très difficile d'assigner à l'hydrologie un but plus précis que la lourde tâche de décrire et comprendre le cycle de l'eau, dans un environnement éminemment complexe, hétérogène et variable dans le temps. Le moteur de cette science réside principalement dans le lien étroit qui existe entre l'homme et son environnement, et en particulier dans la dépendance qu'il a vis-à-vis de la ressource en eau, pour ses besoins d'alimentation ou dans ses activités.

La modélisation hydrologique est une discipline des sciences hydrologiques qui s'intéresse principalement au cycle hydrologique afin d'aboutir à un outil de travail exploitable dans différents domaines. Au cours de ces dernières années, les efforts de la recherche pour la compréhension du cycle de l'eau dans les milieux naturels (bassins versants, rivières, nappes, ...etc.), associés aux développements de l'informatique, se sont concrétisés par l'apparition d'une multitude de modèles mathématiques.

Le présent travail est divisé en deux parties et le polycopié sera donc articulé autour de trois chapitres :

PARTIE 1 : Analyse hydrologique

1. Les notions de probabilités et d'analyse statistique appliquées à l'hydrologie seront décrites dans le premier chapitre.
2. Les notions élémentaires sur les corrélations des séries hydrologiques ainsi que l'analyse des données seront présentées dans le second chapitre.

PARTIE 2 : modélisation hydrologique

3. Le troisième chapitre sera consacré à la modélisation hydrologique. Les modèles pluie-débit ; qui représentent la transformation de la pluie en écoulement ; seront étudiés.

A la fin de chaque chapitre, des exercices et des cas illustratifs sont proposés avec des réponses et des solutions.

En vue d'aider les étudiants à maîtriser les concepts clés en analyse et modélisation hydrologique et à réussir leurs évaluations, une panoplie de sujets d'examens est mise à leur disposition à la fin de ce polycopié de cours.

Chapitre 1 : Notions de probabilité et d'analyse statistique appliquée à l'hydrologie

Objectif du chapitre

L'analyse statistique permet de synthétiser l'information hydrologique représentée par des séries de mesure sur plusieurs années en quelques paramètres qui reflètent le phénomène étudié. L'analyse statistique consiste en la formalisation des données observées par une expression mathématique. Le problème consiste à choisir le modèle probabiliste qui représentera au mieux la série expérimentale. C'est l'ajustement théorique.

1. Rappel : Terminologie et notions fondamentales

1.1 Vocabulaire

- Population : l'ensemble de toutes les observations dont on tire les échantillons
- Echantillon : observations X_1, X_2, \dots, X_n d'une variable aléatoire
- Événement : l'occurrence d'une valeur spécifique d'une variable aléatoire.

1.1.1 Les variables

On appelle variable une **caractéristique commune** à l'ensemble des individus d'une étude. La valeur de cette caractéristique varie entre les individus.

Les types de variables

On distingue les **variables qualitatives et quantitatives**

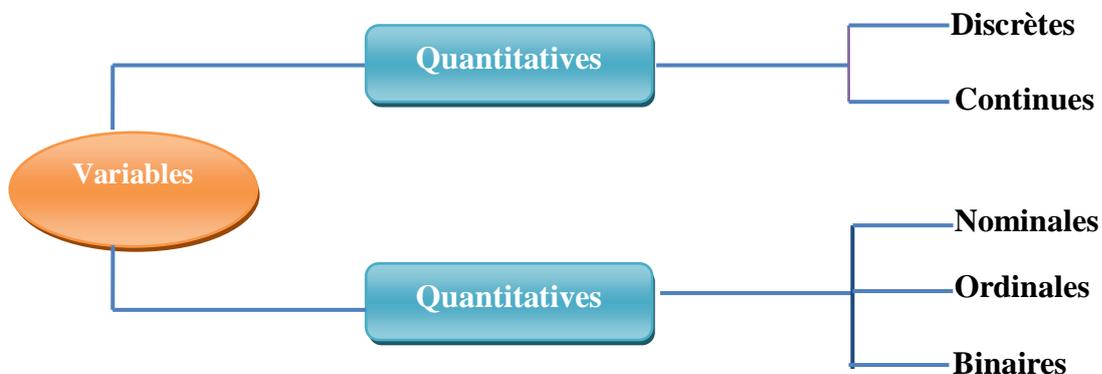


Tableau 1.1 : particularités des variables qualitatives

VARIABLES QUALITATIVES	TYPES	
<ul style="list-style-type: none"> • Elles ne s'expriment pas par une valeur numérique • Elles s'expriment par une qualité 	NOMINALES	<ul style="list-style-type: none"> • Elles ne peuvent pas être hiérarchisées. Aucune valeur n'est supérieure à une autre.

<ul style="list-style-type: none"> • Il est possible de les répartir en classes ou catégories. • On ne peut pas les exploiter mathématiquement. • Si on ne peut pas faire de calcul, on peut cependant les dénombrer, c'est-à-dire compter combien de fois une valeur apparaît. 	ORDINALES	<ul style="list-style-type: none"> • Elles peuvent être classées les unes par rapport aux autres. Il existe un gradient qui permet de les classer dans un ordre logique selon une échelle de valeur.
	BINAIRES	<ul style="list-style-type: none"> • Elles ne peuvent pas être hiérarchisées • Elles ne possèdent que deux modalités (deux classes) possibles.

Tableau 1.2 : Particularités des variables Quantitatives

VARIABLES QUANTITATIVES	TYPES	
<ul style="list-style-type: none"> • S'expriment par un nombre • Peuvent être exploités mathématiquement 	DISCRETES	<ul style="list-style-type: none"> • Elles ne peuvent prendre que des valeurs numériques entières. • Elles se distinguent par le fait que le passage d'une valeur à une autre est discontinu.
	CONTINUES	<ul style="list-style-type: none"> • Elles peuvent prendre n'importe quelle valeur numérique entière ou décimale. • Elles peuvent avoir une infinité de décimales.

1.1.2 Distribution d'une série statistique

En statistique, la distribution statistique, distribution empirique ou distribution des fréquences, est un tableau qui associe des classes de valeurs obtenues lors d'une expérience à leurs fréquences d'apparition. Ce tableau de valeurs est modélisé en théorie des probabilités par une loi de probabilité.

Soit un échantillon de n observations décrivant une variable aléatoire X ($x_1, x_2, x_3, \dots, x_n$).

- La probabilité d'une variable discrète est : $P(X=x) = p(x)$

Elle est caractérisée par : $p(x) \leq 1$

$$F(x) = p(X \leq x) \quad 1.1$$

- Lorsque la variable aléatoire est continue, on parle d'histogramme de fréquences ou polygone de fréquences.

1.1.3 La fonction de densité de probabilité

Si la taille de l'échantillon devient grande et l'intervalle de classe tend vers zéro, le polygone des fréquences relatives sera décrit par une courbe à laquelle est associée une fonction de distribution continue appelée fonction de densité de probabilité. Elle est notée $f(x)$ et caractérisée par :

$$\int_{-\infty}^{+\infty} f(x)dx = 1 \quad 1.2$$

Ainsi l'effectif où la fréquence d'apparition d'une valeur x_i deviendra la densité de probabilité $f(x_i)$.

La fonction de répartition représente la probabilité de non dépassement :

$$p(X \leq x) = \int_{-\infty}^x f(x)dx = F(x) \quad 1.3$$

La probabilité pour que la variable soit comprise entre deux valeurs a et b est :

$$p(a \leq x \leq b) = p(b) - p(a) = \int_a^b f(x)dx \quad 1.4$$

La probabilité au dépassement est :

$$p(X \geq x) = \int_x^{+\infty} f(x) dx = 1 - p(X \leq x) \quad 1.5$$

En hydrologie on parle surtout de probabilité de dépassement (ou probabilité d'apparition) :

$$p(X \geq x) = 1 - F(x) \quad 1.6$$

Exemple :

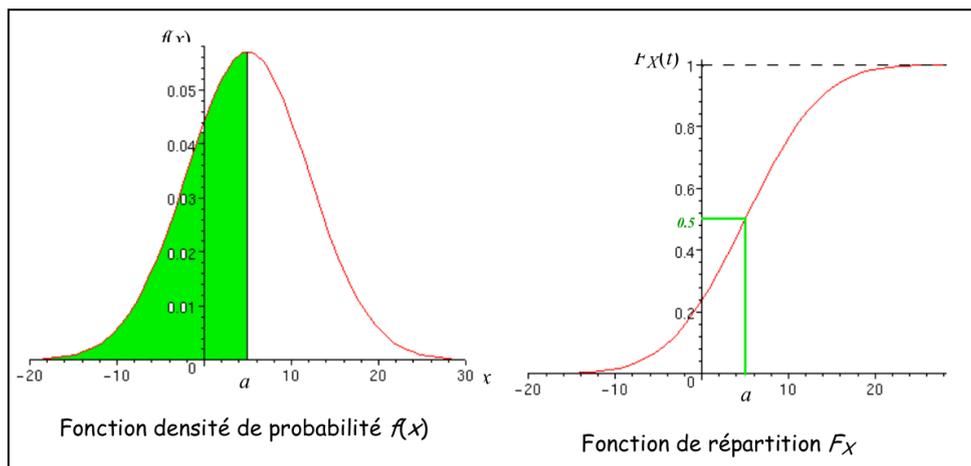


Figure 1.1 : Fonction de densité de probabilité et fonction de répartition

L'aire **hachurée en vert** sous la courbe de la fonction densité de probabilité correspond à la probabilité $P(X < a)$ et vaut **0,5** car ceci correspond exactement à la moitié de l'aire totale sous la courbe. Cette probabilité correspond à la valeur de la fonction de répartition au **point d'inflexion de la courbe**.

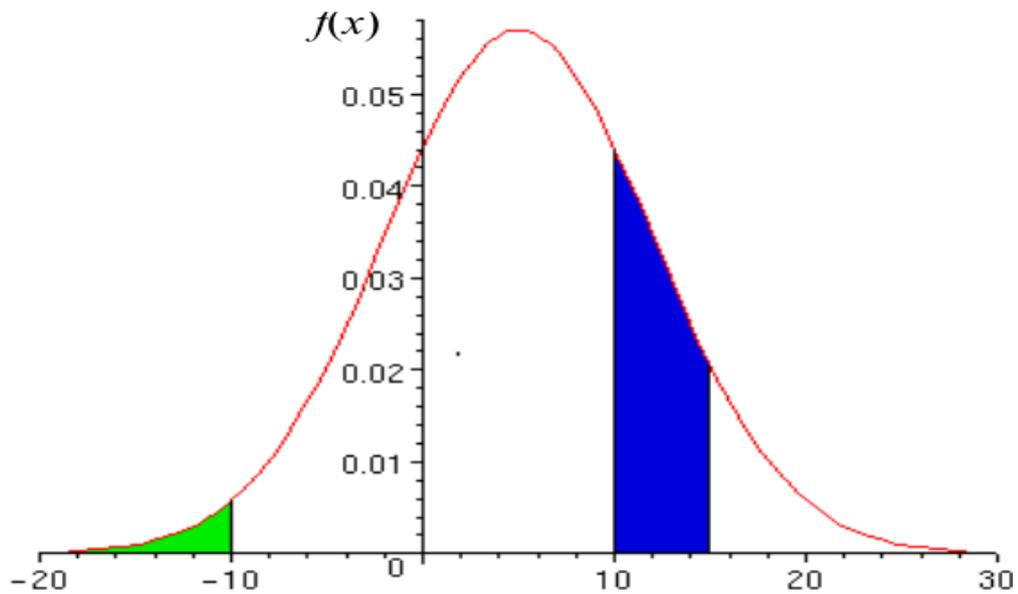


Figure 1.2 : Fonction de densité de probabilité

- L'aire hachurée **en vert** correspond à la probabilité $P(X < -10)$
- l'aire hachurée **en bleu** correspond à la probabilité $P(+10 < X < +15)$

1.1.4 Période de retour

La probabilité d'apparition d'un phénomène est une notion importante pour le dimensionnement des structures conditionnées par un phénomène naturel. On définit l'intervalle moyen de récurrence, ou période de retour, par:

$$T = \frac{1}{p} \quad 1.7$$

P : probabilité de dépassement ($p(X \geq x_T)$)

Exemple :

- un débit d'inondation dont la probabilité d'apparition ou de dépassement est de 0.04 est appelé crue de 20 ans ($T=1/0.04$)

Les intervalles de récurrence recommandés pour le dimensionnement de certaines structures sont présentés dans le tableau suivant.

Tableau 1.3 : Périodes de retour recommandées pour certains ouvrages

Type d'ouvrage	Période de retour (T) recommandée
Déversoirs de barrage	500 à 1000 ans
Ponts	
Autoroutes	100 ans

Routes principales	50 ans
Routes secondaires	25 ans
Digues	100 ans
Plaines inondables	100 ans
Egouts pluvieux, fossés de drainage	5 à 10 ans
Egouts pluvieux de moindre importance	1 à 2 ans

1.1.5 Notions de risque

Le concept de risque hydrologique est à la base du choix de la période de récurrence utilisée pour la conception d'ouvrages hydrauliques. Elle représente la probabilité qu'un critère de conception soit dépassé au moins une fois pendant la période de retour calculée (T) ;

La probabilité que la variable aléatoire prenne la valeur x (probabilité pour qu'un événement se produise) pour la première fois dans les (k-1) prochaines années :

C'est le produit des probabilités de non apparition pendant (k-1) années et la probabilité d'apparition à la (k^{ème}) année :

$$p(1 - p)^{k-1} = \frac{1}{T} \left(1 - \frac{1}{T}\right)^{k-1} \quad 1.8$$

On définit le risque hydrologique (R) comme étant la probabilité de dépassement de la valeur x_T au cours des k années de la vie d'un projet. La probabilité P_k que l'événement x se produise au moins une fois dans les (k) prochaines années : C'est la somme des probabilités d'apparition pendant les années 1, 2, 3, ..., k.

$$P_k = p + p(1 - p) + p(1 - p)^2 + p(1 - p)^3 + \dots + p(1 - p)^{k-1} = 1 - (1 - p)^k$$

$$p_k = R = 1 - \left(1 - \frac{1}{T}\right)^k \quad 1.9$$

Exemple 1:

La probabilité d'apparition dans les 40 prochaines années (risque), d'un débit de valeur x dont l'intervalle de récurrence est T=1000 ans:

$$p_{40} = 1 - \left(1 - \frac{1}{1000}\right)^{40} = 0.96$$

Pour un échantillon de 40 ans, il y a 96 sur 100 de chance (de risque) que cet échantillon contienne une valeur dont l'intervalle réel de récurrence est de 1000 ans. En spécifiant le risque, on peut déterminer la période de récurrence nécessaire.

Exemple 2:

Un risque de 10 % pour que la capacité d'un ouvrage ne soit pas dépassée durant les 25 prochaines années :

$$0.01 = 1 - \left(1 - \frac{1}{T}\right)^{25}$$

La période de récurrence nécessaire est $T = 238$ ans.

1.2 Caractéristiques d'une distribution

1.2.1 Paramètres de tendance centrale

1. *Moyenne arithmétique* : La moyenne arithmétique d'une série ou moyenne arithmétique simple se calcule par une formule qui est donnée par l'expression :

$$\bar{x} = \frac{\sum x_i}{N} \quad 1.10$$

2. *La moyenne géométrique* : de n valeurs positives x_i est la racine $n^{\text{ième}}$ du produit de ces valeurs. Notée G , elle s'écrit

$$G(x_1, x_2 \dots \dots, x_n)^{\frac{1}{n}} = \sqrt[n]{x_1 \cdot x_2 \dots \dots x_n} \quad 1.11$$

3. *La médiane* : On ne peut la calculer que pour les caractères quantitatifs. Les valeurs étant classées par ordre croissant, la médiane est la valeur du caractère qui partage celui-ci en deux ensembles d'effectifs égaux : 50 % des valeurs lui sont supérieures et 50 % lui sont inférieures. La médiane est la valeur la plus proche de toutes les autres c'est celle qui minimise les distances en valeur absolue

4. *Le mode* : La valeur la plus fréquente d'une série statistique — C'est la (ou les) valeur(s) du caractère dont l'effectif est le plus grand.

Le mode peut être utilisé pour résumer des variables catégoriques, alors que la moyenne et la médiane ne peuvent être calculées que pour les variables numériques. C'est d'ailleurs le principal avantage de cette mesure. Il est aussi utile pour les variables discrètes et pour les variables continues lorsqu'elles sont présentées par intervalles.

1.2.2 Paramètres de dispersion

L'étendue :

L'étendue d'une distribution est égale à la différence entre la plus grande et la plus petite valeur de la distribution :

$$w = x_{max} - x_{min} \quad 1.12$$

Quantiles

Pour remédier aux inconvénients de l'étendue, on peut retirer les valeurs les plus extrêmes et calculer l'intervalle des valeurs restantes : c'est la base de la méthode des quantiles.

On appelle quantiles les bornes d'une partition en classes d'effectifs égaux. Attention, lorsqu'on utilise les quantiles, ce sont les effectifs qui sont égaux et non pas les amplitudes.

L'intervalle interquartile (Q3-Q1) :

C'est un paramètre de dispersion absolue qui correspond à l'étendue de la distribution une fois que l'on a retiré les 25% des valeurs les plus faibles et les 25% des valeurs les plus fortes. 50% des observations sont donc concentrées entre Q1 et Q3.

1^{ère} méthode pour déterminer les quartiles (utilisation de la médiane)

On remarque que le second quartile (Q2) d'une distribution X est égal à la médiane Med(X) de cette distribution. On détermine cette valeur à l'aide des méthodes exposées précédemment puis on découpe la distribution X en deux demi-distributions X' et X'' correspondant respectivement aux valeurs supérieures ou inférieures à la médiane.

On remarque alors que Q1 est égal à la médiane de la distribution X' et que Q3 est égale à la médiane de la distribution X'', valeurs qui peuvent être obtenues sans difficultés.

Le seul problème de cette méthode concerne l'établissement des deux demi-distributions X' et X'' lorsque le nombre d'observations est impair. La médiane correspondant alors à une valeur du caractère, on devra la placer dans chacune des demi-distributions pour trouver ensuite Q1 et Q3.

2^{ème} méthode pour déterminer les quartiles (utilisation des fréquences cumulées)

Une seconde méthode plus générale pour déterminer n'importe quel quantiles (et pas seulement les quartiles) consiste à utiliser les fréquences cumulées.

Par définition, les fréquences cumulées correspondant aux quartiles sont en effet les suivantes :

$$F_{\text{cum}}(Q1)=25\%$$

$$F_{\text{cum}}(Q2)=50\%(\text{médiane})$$

$$F_{\text{cum}}(Q3)=75\%$$

Il suffit donc de lire sur la courbe des fréquences cumulées les valeurs approximatives correspondant à Q1, Q2 et Q3 ou à procéder à une interpolation linéaire entre les bornes d'une partition en classes pour trouver les valeurs exactes.

Bien que les quartiles soient les plus utilisés, on peut également utiliser d'autres découpages d'effectifs égaux :

- les **quintiles** partagent une distribution en cinq classes d'effectifs égaux et sont très utilisés en cartographie.

- les **déciles** partagent une distribution en dix classes d'effectifs égaux. L'intervalle compris entre le premier décile et le dernier décile est **l'intervalle inter déciles (D9-D1)**. L'intervalle inter déciles est l'amplitude de la distribution une fois retirée les 10% de valeur les plus fortes et les 10% de valeurs les plus faibles.

Ecart absolu moyen / Ecart absolu médian

L'écart absolu moyen (e_a) est la moyenne de la valeur absolue des écarts à la moyenne. Autrement dit, c'est la distance moyenne à la moyenne. Bien qu'il soit moins utilisé, on peut calculer de la même manière l'écart absolu médian (e_p) qui est la moyenne des écarts à la médiane.

$$e_a = \frac{\sum xi - \bar{x}}{N} \quad 1.13$$

$$e_p = \frac{Q3 - Q1}{2} \quad 1.14$$

Ecart-type

L'écart-type (σ) est la racine carrée de la variance (**var**), elle-même définie comme la moyenne du carré des écarts à la moyenne :

$$Var = \frac{1}{N} \sum_{i=1}^n (xi - \bar{x})^2 \quad 1.15$$

$$\sigma = \sqrt{var} \quad 1.16$$

1.3 Analyse fréquentielle

Il s'agit d'un ensemble de méthodes statistiques permettant d'étudier des données historiques ; en hydrologie ce traitement concerne les données de pluie, de débit, etc..

L'objectif de ces méthodes est de déterminer les probabilités d'occurrence d'un événement spécifique et par conséquent les périodes de retour ou les fréquences d'apparition d'un ensemble d'événements. L'évolution des moyens de calculs permet maintenant d'envisager la simulation en continue d'un grand nombre d'événements, et de réaliser l'analyse fréquentielle des différentes conséquences de ces événements.

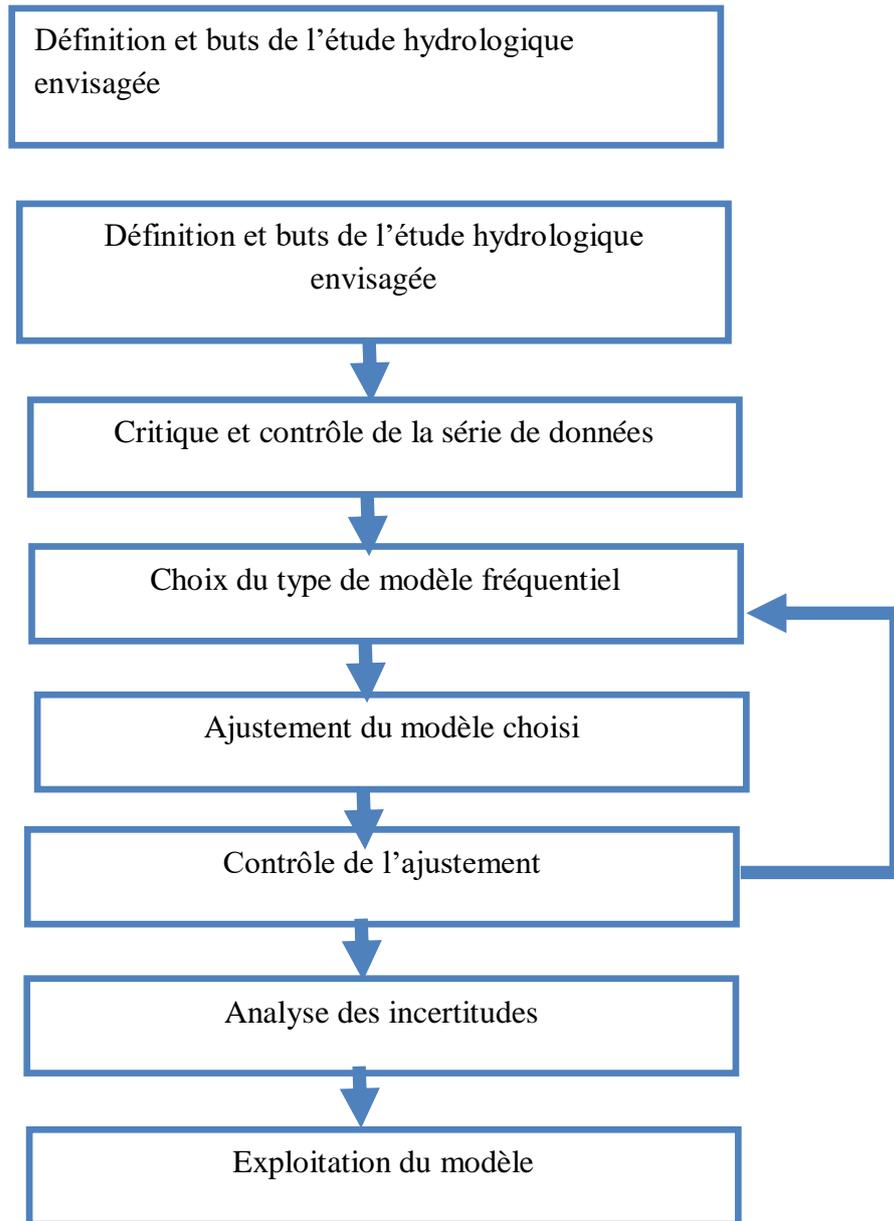


Figure 1.3 : Etapes d'applications de l'analyse fréquentielle

2. Types de fonctions de distribution en hydrologie

Les lois mathématiques de distribution de probabilité peuvent être comparées à la distribution de l'échantillon disponible. Si la loi mathématique s'applique à l'échantillon, on peut déclarer que cette loi s'applique au phénomène et en déduire les probabilités de toutes les valeurs que peut prendre la variable étudiée. Les échantillons étant petits en hydrologie, il sera parfois difficile de déclarer si une loi de probabilité s'applique d'une façon absolue à un phénomène. On devra souvent se contenter d'utiliser la loi qui semble le mieux s'appliquer au phénomène étudié, représentée d'une façon plus ou moins parfaite par l'échantillon disponible. Fonctions de densité des probabilités et de Distribution.

La loi de probabilité doit être susceptible de s'ajuster d'une manière adéquate à l'échantillon. Cette pratique statistique consiste à ajuster une loi théorique de répartition des fréquences connue à la courbe

expérimentale des fréquences de l'échantillon. La loi théorique doit représenter la courbe expérimentale le plus fidèlement possible.

Dans la pratique hydrologique, on choisit à priori une loi théorique (Gauss, Galton, Gumbel, ...) et l'ajustement consiste à déterminer au mieux les paramètres de cette loi. Ensuite, on pourra estimer la valeur d'un évènement quelconque pour une probabilité donnée.

Remarque :

La valeur d'un évènement correspondant à une probabilité d'apparition doit correspondre à une période de retour inférieure à 3 fois la longueur de l'échantillon.

2.1 La loi normale

La fonction de répartition de la loi normale (Gauss) est :

$$F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^u e^{-1/2u^2} du \quad 1.17$$

Avec : $u = \frac{x-\bar{x}}{\sigma}$

\bar{x} : moyenne de la variable aléatoire

σ : écart type

u : variable normale réduite.

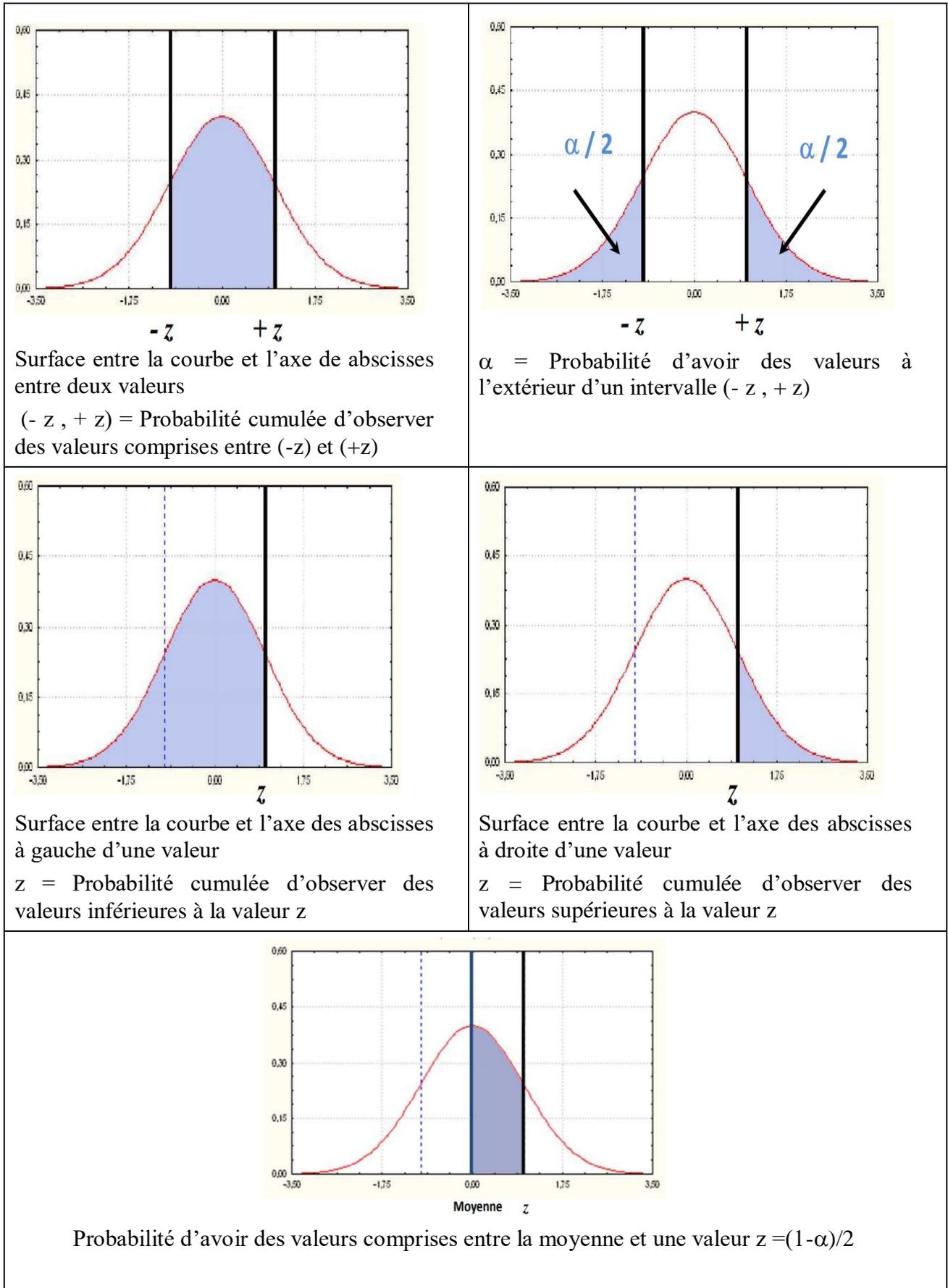
La loi de Gauss offre une répartition symétrique de part et d'autre de la moyenne.

La droite de Henri a pour équation : $x = \sigma u + \bar{x}$ 1.18

Pour chaque valeur x_i de l'échantillon, on peut calculer sa variable réduite u_i . La table de Gauss donne la probabilité de non dépassement de u_i (annexe 01) et par conséquent la probabilité de non dépassement de la valeur x_i .

Remarque :

Il existe plusieurs types de la table de Gauss et ceci en fonction des aires limitées par la courbe $N(0,1)$ (Figure 1.4).



Surface entre la courbe et l'axe de abscisses entre deux valeurs
 $(-z, +z)$ = Probabilité cumulée d'observer des valeurs comprises entre $(-z)$ et $(+z)$

α = Probabilité d'avoir des valeurs à l'extérieur d'un intervalle $(-z, +z)$

Surface entre la courbe et l'axe des abscisses à gauche d'une valeur
 z = Probabilité cumulée d'observer des valeurs inférieures à la valeur z

Surface entre la courbe et l'axe des abscisses à droite d'une valeur
 z = Probabilité cumulée d'observer des valeurs supérieures à la valeur z

Probabilité d'avoir des valeurs comprises entre la moyenne et une valeur $z=(1-\alpha)/2$

Figure 1.4 : différents types de la table de Gauss et ceci en fonction des aires limitées par la courbe $N(0,1)$

Les étapes d'application de la loi normale

On dispose d'une série de N observations. On se propose d'ajuster la loi de Gauss à la répartition expérimentale des fréquences. Les modalités à suivre sont :

- Calcul des caractéristiques empiriques :

- La moyenne : $\bar{x} = \frac{\sum x_i}{N}$

- L'écart type : $\sigma = \sqrt{\frac{1}{N} \sum x_i^2 - N\bar{x}^2}$

-Le coefficient de variation : $CV = \frac{\sigma}{\bar{x}}$

- Classement des valeurs par ordre décroissant en attribuant à chacune son numéro d'ordre i à compter à partir de 1.
- Calcul de la fréquence expérimentale de dépassement pour chacune des valeurs suivant la formule : $f_i = \frac{i-0.5}{N}$
- Le report de l'échantillon en graphique, sur du papier à probabilité normale, avec en ordonnées arithmétiques les valeurs observées (xi) et en abscisses les fréquences expérimentales calculées (fi). L'alignement des points permet d'un premier abord à prévoir que la loi de Gauss peut être ajustée à la série.
- Le tracé de la droite de Henri en reportant deux points : la moyenne et une valeur quelconque.

2.2 Loi log-normale (Galton)

La loi de Galton résulte de la loi normale, mais rendue dissymétrique par un changement de variable. On prend comme variable gaussienne, non pas la variable étudiée x mais son logarithme népérien ou une fonction linéaire de ce logarithme.

On représente généralement la loi log-normale sous forme :

$$F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^u e^{-1/2u^2} du \quad 1.19$$

Avec $u = \frac{x - \overline{\ln(x)}}{\sigma_{\ln(x)}}$

L'équation de la variable réduite u présentée sous la forme :

$$\ln(x) = \sigma_{\ln(x)}u + \overline{\ln(x)} \quad 1.20$$

Est l'équation d'une droite tracée sur papier à probabilité log-normale (papier Galton) qui est un système de coordonnées dont l'abscisse est en échelle gaussienne et l'ordonnée est en échelle logarithmique.

Les étapes d'application de la loi log normale

Notions de probabilité et d'analyse statistique appliquée à l'hydrologie

- Classement des valeurs de l'échantillon par ordre décroissant en attribuant à chacune son numéro d'ordre i à compter à partir de 1.
- Calcul de la fréquence expérimentale de dépassement pour chacune des valeurs suivant la formule : $f_i = \frac{i-0.5}{N}$.
- Détermination des logarithmes népériens de toutes les valeurs de l'échantillon.
- Calcul des caractéristiques empiriques de l'échantillon des logarithmes :

- La moyenne $\overline{\ln(x)} = \frac{\sum \ln(x)}{N}$ 1.21

- L'écart type de la série des logarithmes $\sigma_{\ln(x)} = \sqrt{\frac{1}{N} \sum (\ln(x))^2 - N \overline{\ln(x)}^2}$ 1.22

- Le report sur papier à probabilité log_normale les point expérimentaux
- Le tracé de la droite $\ln(x) = \sigma_{\ln(x)} u + \overline{\ln(x)}$ 1.23

Pour estimer la valeur de la variable étudiée de probabilité donnée (x_p), il faut au préalable estimer le logarithme de cette valeur en utilisant l'équation:

$$\ln(x_p) = \sigma_{\ln(x)} u_p + \overline{\ln(x)} \quad 1.24$$

En prenant dans la table de Gauss la valeur u_p correspondant à la probabilité voulue p .

La valeur de la variable (x_p) sera obtenue par extraction de son logarithme :

$$x_p = e^{\sigma_{\ln(x)} u_p + \overline{\ln(x)}} \quad 1.25$$

2.3 Loi de Gumbel

La fonction de répartition de la loi de Gumbel est (Gumbel E, 1958):

$$F(x) = e^{-e^{\alpha(x-x_0)}} \quad 1.26$$

Avec

$F(X)$ = fréquence de non-dépassement de la valeur X

α, x_0 = coefficients d'ajustement.

Par un changement de variable : $y = \alpha(x - x_0)$ 1.27

La loi de Gumbel s'écrit : $F(x) = F(y) = e^{-e^{-y}}$ 1.28

Où y est la variable réduite de Gumbel, liée à la probabilité attachée à la valeur x .

$F(y)$: fréquence de non dépassement de la variable réduite de y .

A partir de l'équation 1.27:

$x = \frac{1}{\alpha} y + x_0$ est l'équation d'une droite qui représente la loi de Gumbel avec :

$\frac{1}{\alpha}$ est la pente de la droite (Gradex) et x_0 est l'ordonnée à l'origine (mode).

La loi de Gumbel est représentée sur du papier à probabilité Gumbel porte en graduation d'abscisses une échelle de fréquence au non dépassement et une échelle arithmétique de de la variable réduite. A chaque valeur de y de la seconde échelle correspond sur la première échelle la valeur de la fréquence au non dépassement calculée selon la formule $F(x) = e^{-e^{-y}}$. L'ordonnée représente une échelle arithmétique de la variable étudiée (x).

La représentation graphique de l'échantillon des valeurs extrêmes permet au premier abord de voir si la loi de Gumbel est ajustable.

Dans le cas d'un alignement satisfaisant, on peut calculer les paramètres de la droite de régression en utilisant la méthode des moindres carrés.

Dans ce cas :

$$\frac{1}{\alpha} = r \frac{\sigma_x}{\sigma_y} = \frac{\sum_{i=1}^N y_i x_i - N \bar{x} \bar{y}}{\sum_{i=1}^N y_i^2 - N \bar{y}^2} \quad 1.29$$

$$x_0 = \bar{y} - \frac{1}{\alpha} \bar{x} \quad 1.30$$

Avec :

- Y_i : variable réduite de Gumbel
- X_i : variable étudiée.

Les valeurs de y_i correspondant à chacune des fréquences expérimentales peuvent être lues directement sur le graphique ou en utilisant deux fois à l'envers la fonction exponentielle e^{-y} . dans ce cas, $y = -[\ln(-\ln F(x))]$. On constitue ainsi une série de N valeurs de y .

- r : coefficient de corrélation linéaire entre la série des x et y .
- σ_x : écart type de la série de la variable étudiée x .
- σ_y : écart type de la série des N variables réduites y .

Les étapes d'application de la loi de Gumbel

La réalisation par la méthode graphique de l'ajustement d'une loi de Gumbel à un échantillon de N valeurs extrêmes nécessite les opérations suivantes :

- Classement des N valeur extrêmes annuelles par ordre de grandeur croissant.
 - Calcul pour chacune des valeurs, sa fréquence expérimentale au non-dépassement
- $$F(X) = \frac{n-1}{N} \quad (1.31)$$
- Report des couple ($x, f(X)$) sur papier diagramme à probabilité Gumbel.
 - Calcul des paramètres d'ajustement x_0 et $\frac{1}{\alpha}$ selon les équations 1.29 et 1.30

- Tracé de la droite d'ajustement $x = \frac{1}{\alpha}y + x_0$. Pour cela, il suffit de prendre 2 valeurs de probabilité pour lesquelles, il faut déterminer la variable y qu'il suffira de remplacer dans la droite de régression.

3. Valeur extrêmes d'une variable en hydrologie (Rodier, 1994)

Pour l'évaluation des ressources en eau de surface d'un pays ou d'une région l'étude du bilan a une très grande importance, et c'est tout à fait normal, mais on passe parfois sous silence la méthodologie d'étude des données extrêmes: sécheresses et crues exceptionnelles, et c'est regrettable, car ces données extrêmes peuvent jouer un très grand rôle dans la conception des projets d'aménagement. Dans bien des cas, il est nécessaire de connaître les modalités de remplissage d'un réservoir artificiel en cas de sécheresse et il existe des barrages pour lesquels l'ouvrage qui évacue la crue du projet coûte plus cher que le reste de l'aménagement, ce qui impose une analyse correcte de la crue exceptionnelle. Au niveau de la planification de l'aménagement des ressources en eau, les données concernant les sécheresses peuvent imposer des priorités absolues: exploitation intégrée des ressources en eaux souterraines et superficielles, priorité en zones arides à l'aménagement de certains grands fleuves dont les apports restent suffisants même en cas de déficits hydriques exceptionnels. Les crues exceptionnelles lorsqu'elles sont particulièrement dévastatrices peuvent transformer les règles d'aménagement de toutes les zones basses.

a. La sécheresse (Rodier, 1994)

Définitions: De façon générale il y a sécheresse lorsqu'il se produit pendant un temps assez long (mois, année ou suite d'années) et sur une superficie notable t (au moins 20.000 - 100.000 km²) un déficit de disponibilités naturelles en eau par rapport à la valeur moyenne de ces disponibilités qui comportent les précipitations, les écoulements superficiels et souterrains.

Les aspects de la sécheresse varient beaucoup suivant le type de ressources affecté par ce phénomène naturel et suivant le type d'activité économique le plus touché : sécheresse agricole, météorologique et la sécheresse hydrologique.

Caractères des sécheresses : Le caractère le plus important est la persistance du déficit qui peut durer plusieurs semaines, plusieurs mois ou plusieurs années.

b. Les crues

Définitions : il existe plusieurs définitions du phénomène hydrologique « Crues à savoir :

Montée importante du niveau d'un fleuve dont les eaux sortent du lit.
Trouvé sur <http://1ber.free.fr/Ensgmnt/FichMeth/Lexique.htm>

Montée des eaux d'un fleuve, d'une rivière.
Trouvé sur <http://g.mace.free.fr/vochist5.htm>

Montée des eaux d'une rivière, généralement assez rapide. De la petite crue de printemps, habituelle et inoffensive, à la crue centennale, voire millénaire, extrêmement destructrice, l'éventail des crues et de leur impact est large. L'intensité et la violence d'une crue sont directement fonction des précipitations qu'a reçues le bassin versant.

Trouvé sur <http://www.glossaire-eau.fr/glossaire-recherche?search=&themes=All&page=54>

Phénomène caractérisé par une montée plus ou moins brutale du niveau d'un cours d'eau, liée à une croissance du débit jusqu'à un niveau maximum. Ce phénomène peut se traduire par un débordement du lit mineur. Les crues font partie du régime d'un cours d'eau.

Trouvé sur http://www.irma-grenoble.com/05documentation/06glossaire_index.php?lettre=C

Ne pas confondre avec l'inondation, la submersion temporaire de terres sous l'eau dont une crue n'est pas forcément l'origine. La crue est une forte augmentation, un accroissement du débit et de la hauteur d'eau en écoulement d'un fleuve, d'une rivière, d'un cours d'eau. Le mot s'utilise fréquemment quand le débordement du lit mineur du cours d'eau commence à provoquer des dommages. Ce débordement provoque une inondation de zones plus ou moins éloignées des rives, en délimitant ainsi une zone inondable. La crue est suivie par une décrue.

Trouvé sur <https://fr.wikipedia.org/wiki/Crue>

Vocabulaire et caractéristiques associées aux crues

Une crue est caractérisée, à un endroit donné, par son hydrogramme (courbe représentant les variations du débit en fonction du temps) qui est la courbe d'évolution du débit en fonction du temps en un point donné d'un réseau ou d'un cours d'eau ; on s'intéresse particulièrement aux hydrogrammes de crue.

Grandeurs caractéristiques des hydrogrammes de crue

Un hydrogramme de crue est défini par différentes grandeurs plus ou moins facilement mesurables (figure 1.5) :

- le débit de pointe qui correspond à la plus forte valeur du débit q_p ;
- le temps de base t_b : durée pendant laquelle le débit est significativement différent du débit habituel, et qui caractérise la durée de la crue ;
- le temps de montée t_p : temps qui s'écoule entre le début d'une montée significative de l'hydrogramme et l'apparition du débit maximum ;
- le volume total V ou le débit moyen : $q_m = V/t_b$

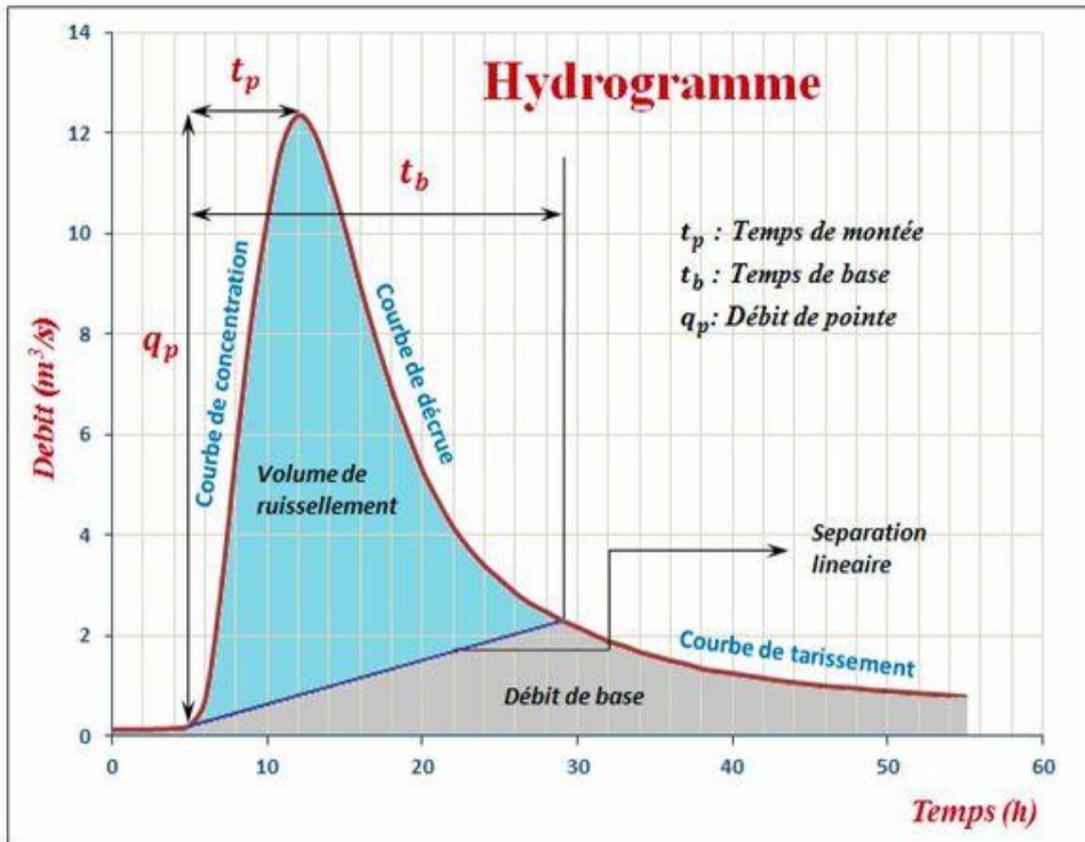


Figure 1.5 : Différentes grandeurs remarquables sur un hydrogramme de crue (Rocha, 2014).

4. Théorie générale de l'ajustement statistique

Les données hydrologiques (débits, précipitations, etc.) sont le résultat de plusieurs processus physiques et sont en général sujettes à toutes sortes d'erreurs. De plus, les séries de données hydrologiques disponibles sont de taille n finie. En pratique, il est alors difficile, et souvent impossible, d'identifier exactement la forme de la loi F de ces observations, c'est à-dire l'ensemble des paramètres qui caractérisent la distribution des probabilités d'occurrence de la variable aléatoire. Ainsi, les moments théoriques et les quantiles d'une variable aléatoire hydrologique demeurent inconnus. Cependant, une estimation des paramètres d'une loi F donnée peut être obtenue à l'aide d'une série finie de réalisations $X_1, X_2, X_3, \dots, X_n$ de la variable aléatoire considérée qui constitue un échantillon. Une estimation des quantiles peut alors en être déduite.

Il existe plusieurs méthodes pour estimer les paramètres d'une loi de probabilité donnée. Nous présentons, tout d'abord, les méthodes bien connues du maximum de vraisemblance et des moments. Par la suite, nous présenterons brièvement la méthode de calcul des intervalles de confiance.

4.1 Estimation par la méthode des moments (Renard, 2014)

L'idée de cette méthode est la suivante : si les paramètres sont bien estimés, alors il devrait y avoir adéquation entre les caractéristiques observées (ou empiriques) et les caractéristiques théoriques. Nous allons rechercher cette adéquation sur les moments, en général le premier moment non centré (c'est à dire la moyenne) et le second moment centré (c'est à dire la variance), plus éventuellement des

moments d'ordre supérieur pour les distributions possédant plus de deux paramètres. Plus précisément, supposons que nous ayons p paramètres $\theta_1, \dots, \theta_p$ à estimer, alors leurs estimateurs $\widehat{\theta}_1, \dots, \widehat{\theta}_p = \widehat{\theta}$ seront solutions du système :

$$\left\{ \begin{array}{l} E_{\widehat{\theta}}(X) = m_1 \\ Var_{\widehat{\theta}}(X) = m'_2 \\ \cdot \\ \cdot \\ E_{\widehat{\theta}}[(X - E_{\widehat{\theta}}(X))^p] = m'_p \end{array} \right. \quad 1.32$$

La notation $E_{\widehat{\theta}}$ a pour but de bien montrer que les moments théoriques sont des fonctions paramètres à estimer.

Exemple 1 : estimateurs des paramètres d'une loi normale

Soit (x_1, \dots, x_n) un échantillon réalisation d'une loi normale $N(\mu, \sigma)$. L'espérance vaut \bar{x} et la variance σ^2 . Les estimateurs des moments sont définis par :

$$\widehat{\mu} = m_1 = \bar{x} \quad 1.33$$

$$\widehat{\sigma}^2 = m_2 = Var_x \Rightarrow \widehat{\sigma} = S_x \quad 1.34$$

En d'autres termes, les estimateurs des moments sont tout simplement égaux à la moyenne et à l'écart-type empiriques.

4.2 Estimation par la méthode du maximum de vraisemblance (Renard, 2014)

Cette méthode consiste de choisir comme estimateur $\theta_1, \dots, \theta_p$ les valeurs qui rendent l'échantillon le plus probable possible. Nous appellerons vraisemblance de l'échantillon (x_1, \dots, x_n) la fonction :

$$L(x_1, \dots, x_n | \theta) = \prod_{i=1}^n f(x_i | \theta) \quad 1.35$$

La vraisemblance d'un échantillon est le produit des densités de probabilités calculées en chaque observation. Les valeurs (x_1, \dots, x_n) ayant été observées et étant connues, la vraisemblance est vue comme une fonction de paramètres $\theta_1, \dots, \theta_p$ que l'on cherche à estimer. Il faut donc trouver les valeurs $\widehat{\theta}_1, \dots, \widehat{\theta}_p$ qui maximisent cette vraisemblance. Dans la pratique, on aura souvent intérêt à maximiser la **log-vraisemblance**.

$$\log L(x_1, \dots, x_n | \theta) = \sum_{i=1}^n \log(f(x_i | \theta)) \quad 1.36$$

Pour ce faire, il faut résoudre p équations aux dérivées partielles :

$$\frac{\partial \log L}{\partial \theta_i} = 0, \forall i = 1, \dots, p \quad 1.37$$

Exemple : Estimation du paramètre d'une loi exponentielle $\text{Exp}(0, \lambda)$

Soient (x_1, \dots, x_n) des observations issues de cette loi exponentielle, dont la densité s'écrit :

$$f(x|\lambda) = \frac{1}{\lambda} e^{-x/\lambda} \quad 1.38$$

La log vraisemblance vaut donc :

$$\log L(x_1, \dots, x_n | \lambda) = \sum_{i=1}^n \log(f(x_i | \lambda)) \quad 1.39$$

$$= \sum_{i=1}^n \log\left(\frac{1}{\lambda} e^{-\frac{x_i}{\lambda}}\right) \quad 1.40$$

$$= n \log\left(\frac{1}{\lambda}\right) - \sum_{i=1}^n \frac{x_i}{\lambda} \quad 1.41$$

$$= -n \log(\lambda) - \sum_{i=1}^n \frac{x_i}{\lambda} \quad 1.42$$

D'où :

$$\frac{\partial \log L}{\partial \lambda} = -\frac{n}{\lambda} + \sum_{i=1}^n \frac{x_i}{\lambda^2} = 0 \quad 1.43$$

$$\Leftrightarrow -n + \sum_{i=1}^n \frac{x_i}{\lambda} = 0 \quad 1.44$$

$$\Leftrightarrow \lambda = \frac{1}{n} \sum_{i=1}^n x_i \quad 1.45$$

L'estimateur du maximum de vraisemblance correspond ici simplement à la moyenne empirique des observations.

5. Intervalles de confiance et bandes de confiance (Serra Bertral, 1979)

La connaissance de la seule valeur estimée d'un paramètre empirique ne présente guère d'intérêt si on n'a aucune précision de l'estimation obtenue. Il est nécessaire de compléter l'information calculée, en déterminant l'intervalle qui contient la vraie valeur du paramètre. Cette manière de procéder conduit à la notion d'intervalle de confiance.

Supposons que l'on s'intéresse à un paramètre quelconque d'une population. Par exemple la moyenne d'un échantillon \bar{x} . on se propose de déterminer d'une part et d'autre de \bar{x} les limites $x_1 < \bar{x} < x_2$ qui ont une forte probabilité de contenir la vraie « valeur » m .

On détermine les limites de confiance x_1 et x_2 de telle sorte que :

$$\text{Prob}(x_1 < \bar{x} < x_2) = \alpha \quad 1.46$$

On appelle α le coefficient de confiance ou coefficient de sécurité. On choisit arbitrairement le coefficient de sécurité voisin de 1 (0.95 ou 0.99) suivant le risque que l'on accepte de prendre.

L'intervalle x_1 et x_2 est appelé intervalle de confiance. En affirmant que cet intervalle (x_1, x_2) contient la vraie valeur m , on peut commettre une erreur qui est précisément $1-\alpha$. La fixation du risque impose :

$$\text{Prob}(m < x_1 \text{ ou } m > x_2) = 1 - \alpha \quad 1.47$$

Ou

$$\text{Prob}(m < x_1) + \text{Prob}(m > x_2) = 1 - \alpha \quad 1.48$$

Généralement, on divise le risque d'erreur en deux parties égales, les limites de l'intervalle de confiance étant telles que :

$$Prob(m < x_1) = Prob(m > x_2) = 1 - \frac{\alpha}{2} \quad 1.49$$

Calcul de l'intervalle de confiance de la moyenne :

a) La moyenne empirique \bar{x} d'un échantillon de longueur N (qui suit une loi normale) est également une variable normale dont :

- La moyenne est égale à \bar{x}
 - La variance est égale à $\frac{\sigma^2}{N}$
 - L'écart type est égal à $\frac{\sigma}{\sqrt{N}}$
- } Valable pour $N \geq 30$ 1.50

L'écart type est égal à $\frac{\sigma}{\sqrt{N}}$ est appelé l'erreur type sur la moyenne.

b) La variance calculée sur l'échantillon est une variable aléatoire qui suit une loi normale :

- La moyenne est égale à σ^2
 - La variance est égale à $\frac{\sigma^2}{2N}$
 - L'écart type est égal à $\frac{\sigma}{\sqrt{2N}}$
- } Valable pour $N \geq 50$ 1.51

L'écart type est égal à $\frac{\sigma}{\sqrt{2N}}$ est appelé l'erreur type sur la variance.

Pour un grand échantillon ($N \geq 30$ pour la moyenne et $N \geq 50$ pour la variance), pour un degré de confiance α , les limites de l'intervalle de confiance seront :

Pour la moyenne :

$$\bar{x}_{1,2} = \bar{x} \pm u_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{N}} \quad 1.52$$

Pour l'écart type :

$$\bar{\sigma}_{1,2} = \sigma \pm u_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{2N}} \quad 1.53$$

$u_{1-\frac{\alpha}{2}}$ est la variable réduite de Gauss correspondant à la probabilité $1 - \frac{\alpha}{2}$

Remarque :

Dans le cas d'un petit échantillon $N < 30$, la moyenne empirique suit une loi de Student à N-1 degrés de liberté (loi symétrique qui se confond avec la loi de Gauss pour $N \geq 30$). Dans ce cas pour un degré de confiance $1-\alpha$, les limites de l'intervalle de confiance sont :

- Pour $N < 30$ $\bar{x}_{1,2} = \bar{x} \pm t_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{N}}$ 1.54

t est la variable de Student pour N-1 de degrés de liberté (annexe 02).

- Pour $N < 50$, la variance empirique σ^2 suit une de χ^2 à $N-1$ degrés de liberté. Les limites de l'intervalle de confiance, avec un degré de confiance α sont :

$$\frac{N\sigma^2}{\chi^2_1} < \sigma^2 < \frac{N\sigma^2}{\chi^2_2} \quad 1.55$$

Avec :

$$\text{Prob}(\chi^2 < \chi^2_1) = \text{Prob}(\chi^2 < \chi^2_2) = 1 - \frac{\alpha}{2} \quad 1.56$$

Calcul de l'intervalle de confiance pour les valeurs extrêmes

Si la loi de Gauss est ajustable à un échantillon, la valeur que prendrait la variable pour une probabilité « p » peut être déterminée à l'aide de la droite de Henri :

$$x_p = \bar{x} + \sigma u_p \quad 1.57$$

Les bornes de l'intervalle de confiance à $\alpha\%$ du quantile x_p sont:

$$x_{p1,2} = x_p \pm u_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{2N}} \cdot \sqrt{2 + u_p^2} \quad 1.58$$

6. Tests d'ajustement

Lorsque l'on a procédé à un ajustement d'une loi théorique de probabilité à un échantillon, la question se pose de savoir si cette loi s'adapte de façon satisfaisante. Un examen graphique, pour intéressant qu'il soit, ne saurait suffire. Il faut tester par le calcul la qualité de l'ajustement réalisé.

6.1 Rappels sur les principes des tests statistiques

Une *hypothèse statistique* est simplement une assertion au sujet d'une population que l'on peut mettre à l'épreuve en tirant un échantillon au hasard.

Une utilisation courante des statistiques est la notion de *test*, un test est un mécanisme qui permet de trancher entre deux hypothèses au vu des résultats d'un échantillon : soient H_0 (*l'hypothèse nulle*) et H_1 (*l'hypothèse alternative*) ces deux hypothèses, dont une et une seule est vraie.

Dans le processus de prise de décision, on court le risque de commettre deux sortes d'erreurs distinctes. La première est de rejeter à tort H_0 qui est vraie, c'est *l'erreur de première espèce* et sa probabilité est α , le *seuil d'erreur de test*. La deuxième est d'accepter H_0 alors qu'elle est fautive, c'est *l'erreur de deuxième espèce* et sa probabilité est notée β . Le tableau suivant résume le dilemme du test d'hypothèses : l'état de la réalité est inconnu.

On ne sait si H_0 est vraie ou fautive. Si une décision de rejet ou de non-rejet doit être prise face à cette incertitude, on doit prendre le risque de commettre l'une ou l'autre des erreurs. Il y a quatre résultats possibles d'un test d'hypothèses.

Résultat	Décision	
	H_0 acceptable	H_0 rejetée

Si H_0 est vraie	Décision correcte. Probabilité = $1 - \alpha$ = seuil de confiance	Erreur de 1ère espèce. Probabilité = α = seuil du test
Si H_0 est fausse (H_1 est vraie)	Erreur 2^{ème} espèce. Probabilité = β	Décision correcte. Probabilité = $1 - \beta$ = puissance du test

Plus α sera grand (respectivement petit), plus β sera petit (respectivement grand). Le fait d'imposer un α faible conduit à une règle de décision plus stricte qui aboutit le plus souvent à n'abandonner l'hypothèse H_0 que dans des cas rarissimes, et donc à conserver cette hypothèse quelque fois à tort. Le compromis entre les valeurs de α et β est donc souhaitable bien que difficile à réaliser. Les valeurs les plus courantes sont 5%, 1% ou 10%.

α étant fixé, il faut choisir *une variable de décision*, variable qui doit apporter de l'information sur le problème posé, à savoir le choix entre les deux hypothèses. La loi de cette variable doit être parfaitement connue dans au moins une hypothèse (le plus souvent H_0) afin de ne pas introduire de nouvelles inconnues dans le problème. On appelle alors *région critique* l'ensemble des valeurs de la variable de décision qui conduisent à écarter H_0 au profit de H_1 . On appelle *région d'acceptation* la région complémentaire de la région critique. La zone ou région d'acceptation correspond à l'intervalle dans lequel les différences observées entre les réalisations et la théorie sont attribuables aux fluctuations d'échantillonnage. La région critique ou zone de rejet correspond donc aux intervalles dans lesquels les différences sont trop grandes pour être le fruit du hasard d'échantillonnage.

La construction d'un test est la détermination *a priori* de la région critique sans connaître le résultat de l'expérience. On peut donc résumer cette démarche de la manière suivante:

- Choix de H_0 et de H_1 .
- Choix de la valeur du risque de première espèce α
- Détermination de la région critique en fonction de α
- Calcul de la valeur expérimentale de la statistique
- Rejet ou acceptation de l'hypothèse nulle en fonction de la valeur observée.

Choix du test

Un test paramétrique requiert un modèle à fortes contraintes (normalité des distributions, égalité des variances) pour lequel les mesures doivent être réalisées dans une échelle au moins d'intervalle. Ces hypothèses sont d'autant plus difficiles à vérifier que les effectifs étudiés sont plus réduits.

Un test non paramétrique est un test dont le modèle ne précise pas les conditions que doivent remplir les paramètres de la population dont été extrait l'échantillon.

Dans les deux cas, les données considérées doivent être aléatoire, identiquement distribuées et indépendantes les unes des autres.

Le test qui fournit l'erreur β la plus petite, pour une même valeur α , est par définition le plus puissant (celui ayant la plus grande valeur de la puissance de test $1 - \beta$). En effet, il peut détecter les plus petites différences entre les populations sans pour autant augmenter l'erreur de première espèce.

La majorité des tests statistiques repose sur le respect d'un certain nombre de conditions. Selon le degré de respect de ces conditions d'application, la validité des résultats se trouve plus ou moins affectée et elle l'est d'autant plus que le test est moins robuste. Ainsi, la *robustesse* d'un test équivaut à sa tolérance vis-à-vis du respect des conditions.

Si le statisticien dispose de plusieurs tests pour vérifier une hypothèse, il choisira bien sûr le plus puissant et le plus robuste.

Les tests peu puissants augmentent la probabilité de commettre une erreur de deuxième espèce. Cette erreur peut s'avérer particulièrement grave. En effet, en médecine par exemple, une analyse qui classerait comme malade un individu bien portant peut avoir des conséquences aussi graves qu'une analyse qui classerait comme bien portants des individus malades (erreur de première espèce). Dans de tels cas, il y a intérêt à tracer la *courbe de puissance du test*, qui indique la probabilité de prendre une bonne décision si H_1 est vraie. La puissance est mesurée par la valeur de $1 - \beta$ pour un α donné.

Dans notre étude, l'hypothèse H_0 correspondra à une "stationnarité" de la série et l'hypothèse alternative H_1 à "un changement (rupture ou tendance)".

6.2 Test chi-carré de K. Pearson

Le khi-deux est une statistique permettant de comparer les effectifs (fréquences) observés dans un échantillon avec des fréquences théoriques qui découlent des hypothèses statistiques. Nous ne ferons ici qu'étudier le mode d'emploi de ce test :

H_0 : La population suit la distribution « x »

H_1 : la population ne suit pas la distribution « x »

L'idée est de découper le domaine de la distribution en intervalles. Dans chaque intervalle, on calcule à partir de la loi spécifiée sous H_0 la fréquence théorique attendue. On compte ensuite combien d'observations l'on retrouve dans chaque intervalle. Il suffit alors de comparer les fréquences observées aux fréquences théoriques.

Si :

n_i : le nombre de valeurs observées contenues dans la classe i ;

v_i : le nombre de valeurs qui sur un échantillon est affecté par la loi théorique à la classe i .

Le χ^2 est défini par la relation :

$$\chi^2 = \frac{\sum_{i=1}^k (n_i - v_i)^2}{v_i} \quad 1.59$$

k = nombre de classes choisies.

χ^2 est une variable aléatoire qui peut varier de 0 à l'infini.

Notions de probabilité et d'analyse statistique appliquée à l'hydrologie

- si $\chi^2=0$ donc $n_i = v_i$. La loi théorique suit exactement la répartition expérimentale des fréquences. L'ajustement est parfait.
- Plus la valeur de χ^2 est élevée, plus l'ajustement est discutable.

La quantité χ^2 est une variable aléatoire dont la distribution a été étudiée par Pearson. Les valeurs de sa probabilité dus au seul fait du hasard, sont données par des tables en fonction de valeur propre et du nombre de degré de liberté γ .

$$\gamma = k - 1 - p \quad 1.60$$

Avec :

K= nombre de classe choisies

P= nombre de paramètres de la loi théorique testée sous H0.

Le nombre de liberté est en fait le nombre de paramètres que l'on peut fixer librement dans le phénomène étudié.

On choisi librement le nombre de classe ,k, on peut choisir arbitrairement le nombre ni

Dans le pratique on calculera le X^2 et d'après les tables on déterminera sa probabilité de dépassement.

La table donnée en annexe 03 permet de connaître la valeur numérique du χ^2 qui a $\alpha\%$ de chance d'être dépassé par le seul fait du hasard, en fonction des valeurs de γ .

- Si la probabilité trouvée est supérieure à α , l'ajustement est satisfaisant.
- Si la probabilité trouvée est inférieure à α , la loi choisie doit être rejetée.

Les étapes d'application du test de Khi Carré

1. Les N valeurs expérimentales étant classées par ordre décroissant, on les devise en k classes, de telle sorte que chaque classe contienne au minimum 5 données expérimentales. La classe i est limitée par les valeurs x_i et x_{i+1} .

2. On compte le nombre d'observations n_i dans chaque classe i.

3. On calcule à partir de la loi théorique choisie la valeur v_i

$$v_i = N \int_{x_{i+1}}^{x_i} f(x) dx = N[f_i(x_i) - f_i(x_{i+1})] \quad 1.61$$

Avec

$f_i(x_i)$ = fréquence théorique de dépassement de la borne inférieure de la classe i

$F_i(x_{i+1})$ = fréquence théorique de dépassement de la borne supérieur de la classe i

7. On calcule pour chaque classe la quantité $\frac{(n_i - v_i)^2}{v_i}$ et on fait la somme des k quantités pour obtenir le χ^2 .
8. On calcule le nombre de degré de liberté.

9. Interprétation des résultats.

6.3 Test de Kolmogorov-Smirnov :

Le test de Kolmogorov-Smirnov consiste à mesurer, pour une variable aléatoire continue, la plus grande distance entre la distribution théorique $F_0(x)$ et la distribution expérimentale $F(x)$.

On utilise le test de Kolmogorov Smirnov où

Nous avons donc $H_0 : F(x_i) = F_0(x_i) \quad \forall x_i$ et $H_1 : F(x_i) \neq F_0(x_i) \quad \forall x_i$ pour au moins une valeur de x_i . La distribution empirique, ou observée, se calcule, dans la théorie de Kolmogorov-Smirnov, par la relation classique :

$$D_{obs} = \max(|F(x_i) - F_0(x_i)| \quad \forall x_i) \quad 1.62$$

On rejette H_0 au seuil α si $D_{obs} > D_\alpha$ où :

$$\text{Pour } \alpha = 0.05: D_\alpha = \frac{0.895}{S} \quad 1.63$$

$$\text{Pour } \alpha = 0.01: D_\alpha = \frac{1.035}{S} \quad 1.64$$

$$\text{Avec : } S = \sqrt{N} - 0.01 + \frac{0.85}{\sqrt{N}} \quad 1.65$$

Où N est la taille de l'échantillon.

Les étapes d'application du test de Kolmogorov Smirnov

- a. Tri des données expérimentales par ordre croissant
- b. Calcul des variables centrées réduites u_i
- c. Détermination des fréquences au non dépassement $f_0(x_i)$ correspondant à chaque u_i
- d. Calcul des fréquences expérimentales $f(x_i) = \frac{i}{N}$ 1.66
- e. Calcul des différences :

$$D^+ = |f(x_i) - f_0(x_i)| \quad 1.67$$

$$\text{et } D^- = |f(x_{i-1}) - f_0(x_i)| \quad 1.68$$

- f. Trouver la valeur maximale de D^+ et D^- et la comparer à la valeur critique D_α .



Mise
en
application

Exercice 1

Soient les données d'observations d'une station pluviométrique représentées dans le tableau ci-dessous :

Tableau 1.4 : Séries des données des pluies moyennes annuelles

Années	P _{ann} (mm)	Années	P _{ann} (mm)
1915	351	1926	382,9
1916	469,6	1927	634,4
1917	400,2	1928	512,3
1918	387,6	1929	655
1919	403,7	1930	491,5
1920	559,6	1931	535,6
1921	355,2	1932	319,2
1922	485,1	1933	589,7
1923	360,1	1934	338,6
1924	339,3	1935	466,9
1925	423		

1. Ajuster la série d'observations à la loi normale ;
2. Vérifier l'ajustement en utilisant le test de khi carré et le test de Kolmogorov Smirnov avec $\alpha=5\%$;
3. Calculer la pluie décennale, centennale et millénaire pour la loi normale ;
4. Calculer l'intervalle de confiance de la moyenne et de l'écart type avec $\alpha=5\%$.

Solution

1. L'ajustement à la loi normale :

- Classement de l'échantillon par ordre décroissant
- Calcul de la fréquence expérimentale F_i
- Calcul la moyenne et l'écart type
- Tracé de la droite d'ajustement $x = \sigma u + \bar{x}$

Tableau 1.5 : Ajustement de la série d'observation à la loi normale

Année	x	i	Xclassée	Fi(x)
1915	351	1	655	0,02
1916	469,6	2	634,4	0,07
1917	400,2	3	589,7	0,12
1918	387,6	4	559,6	0,17
1919	403,7	5	535,6	0,21
1920	559,6	6	512,3	0,26
1921	355,2	7	491,5	0,31
1922	485,1	8	485,1	0,36
1923	360,1	9	469,6	0,40
1924	339,3	10	466,9	0,45
1925	423	11	423	0,50
1926	382,9	12	403,7	0,55
1927	634,4	13	400,2	0,60
1928	512,3	14	387,6	0,64
1929	655	15	382,9	0,69
1930	491,5	16	360,1	0,74
1931	535,6	17	355,2	0,79
1932	319,2	18	351	0,83
1933	589,7	19	339,3	0,88
1934	338,6	20	338,6	0,93
1935	466,9	21	319,2	0,98
La moyenne			450,5	
L'écart type			101,187	
Coff de variation			0,22	

Avec :

La moyenne : $\bar{x} = \frac{\sum x_i}{N}$

L'écart type : $\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N x_i^2 - N\bar{x}^2}$

Le coefficient de variation : $Cv = \frac{\sigma}{\bar{x}}$

La droite d'ajustement est donnée par l'équation suivante :

$x=101,18u+450,5$

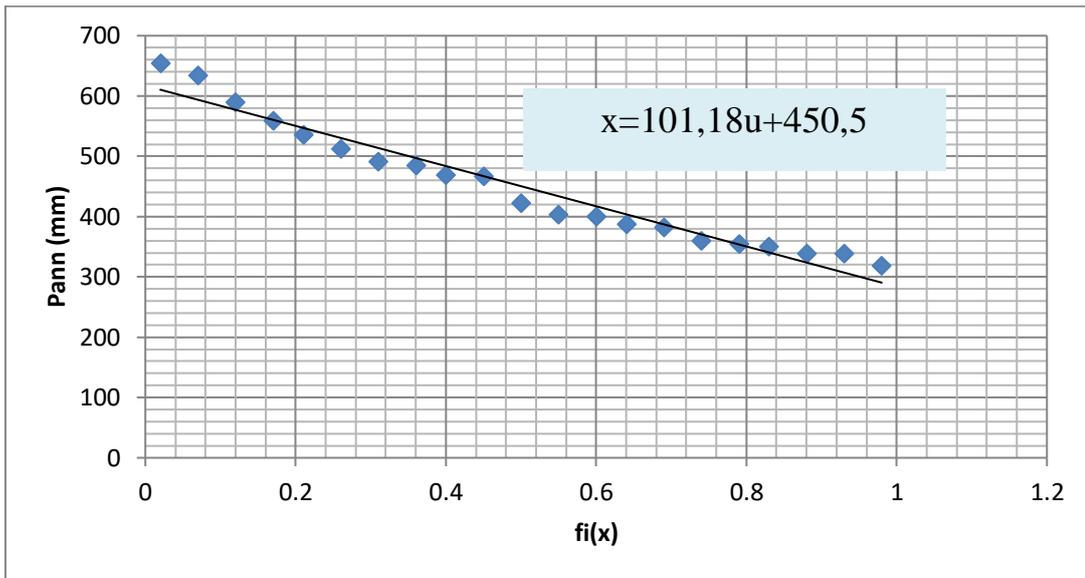


Figure : ajustement graphique à la loi normale

2. Vérification de l'ajustement en utilisant le test de χ^2 :

Les étapes d'application du test de Khi Carré sont répertoriées dans le tableau 1.6 :

Tableau 1.6 : Application du test de khi carré

Classes	ni	Var réduite (Ui)	Variable (Ui+1)	F(ui)	F(Ui+1)	Effectif Théorique (vi)	(ni- vi) ²	(ni- vi) ² / vi
[535,6;+∞[5	0,84	/	0,2005	/	4,21	0,62	0,14
[466,9;535,6[5	0,16	0,84	0,4364	0,2005	4,95	0,00212521	0,000429
[382,9;466,9[5	-0,67	0,16	0,7454	0,4364	6,49	2,21	0,34
] -∞;382,9[6	/	-0,67	1	0,7454	5,35	0,43	0,08

On déduit que :

$\chi^2=0,57$

Interprétation du χ^2 :

Pour interpréter le χ^2 , on calcule le degré de liberté:

$\gamma = K-1-P$

	γ	α	$P(\chi^2)$		Conclusion
$\chi^2=0,57$	1	5%	$0,1 < p(\chi^2) < 0,5$	$P(\chi^2) > 0,05$	Satisfaisant

3. Test de kolmogorov_ Smirnov :

Afin de vérifier l'ajustement des pluies moyennes annuelles à la loi normales, le test de Kolmogorov Smirnov a été utilisé (Tableau 1.7).

Tableau 1.7 : application du test de Kolmogorov Smirnov

Année	x	i	xi (classée)	f(xi)	Ui	f ₀ (x _i)	Di ⁺	Di ⁻
1915	351	1	319,2	0,05	-1,2977	0,0975	0,04988	0,09750
1916	469,6	2	338,6	0,10	-1,1059	0,1357	0,04046	0,08808
1917	400,2	3	339,3	0,14	-1,0990	0,1379	0,00496	0,04266
1918	387,6	4	351	0,19	-0,9834	0,1635	0,02698	0,02064
1919	403,7	5	355,2	0,24	-0,9419	0,1736	0,06450	0,01688
1920	559,6	6	360,1	0,29	-0,8935	0,1867	0,09901	0,05140
1921	355,2	7	382,9	0,33	-0,6681	0,2546	0,07873	0,03111
1922	485,1	8	387,6	0,38	-0,6217	0,2676	0,11335	0,06573
1923	360,1	9	400,2	0,43	-0,4971	0,3121	0,11647	0,06885
1924	339,3	10	403,7	0,48	-0,4625	0,3228	0,15339	0,10577
1925	423	11	423	0,52	-0,2718	0,3936	0,13021	0,08259
1926	382,9	12	466,9	0,57	0,1621	0,5636	0,00783	0,03979
1927	634,4	13	469,6	0,62	0,1888	0,5714	0,04765	-0,00003
1928	512,3	14	485,1	0,67	0,3420	0,6331	0,03357	0,01405
1929	655	15	491,5	0,71	0,4052	0,6554	0,05889	0,01127
1930	491,5	16	512,3	0,76	0,6108	0,7291	0,03280	0,01481
1931	535,6	17	535,6	0,81	0,8411	0,7995	0,01002	0,03760
1932	319,2	18	559,6	0,86	1,0783	0,8577	0,00056	0,04818
1933	589,7	19	589,7	0,90	1,3758	0,9147	0,00994	0,05756
1934	338,6	20	634,4	0,95	1,8176	0,9649	0,01252	0,06014
1935	466,9	21	655	1,00	2,0212	0,9783	0,02170	0,02592
						D _{max}	0,15339	

La valeur maximale :

$$D_{\max} = 0,15339$$

Signification du D_{max} :

Pour $\alpha = 0,05$: $D\alpha = \frac{0,895}{s}$

$$s = \sqrt{N} - 0,01 + \frac{0,85}{\sqrt{N}}$$

$$s = \sqrt{21} - 0,01 + \frac{0,85}{21} = 4,61$$

On déduit que :

$$D\alpha = 0,19$$

$$D_{\max} < D_{\alpha} \Rightarrow 0,15339 < 0,19 \Rightarrow \underline{\text{Ajustement satisfaisant}}$$

4. Calcul de la pluie décennale, centennale et millénaire pour loi normale :

La droite d'ajustement :

$$x = 101,18u + 450,5$$

La pluie décennale :

$$T = 10 \text{ans} \longrightarrow FD = 0,1 \longrightarrow FND = 1 - 0,1 = 0,9$$

$$U = 1,28$$

$$X_{0,1} = 580,01 \text{ mm}$$

La pluie centennale :

$$T = 100 \text{ans} \longrightarrow FD = 0,01 \longrightarrow FND = 1 - 0,01 = 0,99$$

$$U = 2,33$$

$$X_{0,01} = 686,24 \text{ mm}$$

La pluie millénaire :

$$T = 1000 \text{ans} \longrightarrow FD = 0,001 \longrightarrow FND = 1 - 0,001 = 0,999$$

$$U = 3,1$$

$$X_{0,001} = 764,15 \text{ mm}$$

5. Calcul de l'intervalle de confiance de la moyenne :

$$\bar{X} = 450,5 \quad , \quad \partial = 101,18 \quad , \quad N = 21 \quad , \quad \alpha = 5 \%$$

$$\alpha = 0,05 \longrightarrow U_{1-\frac{0,05}{2}} = U_{0,9750} = 1,96$$

$$x_{1.2} = \bar{x} \pm u_{1-\frac{\alpha}{2}} \frac{\partial}{\sqrt{N}}$$

$$X_1 = 450,5 - 1,96 * \frac{101,18}{\sqrt{21}}$$

$$X_2 = 450,5 + 1,96 * \frac{101,18}{\sqrt{21}}$$

On déduit que : **470,22 ≤ \bar{x} ≤ 493,77**

L'erreur type sur la moyenne est: $\frac{\sigma}{\sqrt{N}} = 22.7 \text{ mm}$

Calcul de l'intervalle de confiance de la moyenne de l'écart type :

$$\bar{\sigma}_{1,2} = \sigma \pm u_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{2N}}$$

$$\sigma_1 = 101,18 - 1,96 * \frac{101,18}{\sqrt{2*21}}$$

$$\sigma_2 = 101,18 + 1,96 * \frac{101,18}{\sqrt{2*21}}$$

On déduit que :

$$70,58 \leq \sigma \leq 131,78$$

L'erreur type sur l'écarttype est : $\frac{\partial}{\sqrt{2N}} = 15.61$

Exercice 2

Soient les données d'observations des pluies journalières maximales d'une station pluviométrique représentées dans le tableau ci-dessous :

Tableau 1.8 : Pluies journalières maximales

Années	Pjmax (mm)	Années	Pjmax (mm)
1995	526,2	2006	296,1
1996	682,5	2007	787,3
1997	593,6	2008	502,1
1998	746	2009	513,7
1999	318,8	2010	367,4
2000	524,3	2011	363,9
2001	389,8	2012	301,4
2002	486,3	2013	286,9
2003	396,5	2014	334,9
2004	504,4	2015	1038,7
2005	299		

1. Ajuster la série d'observations à la loi log normale ;
2. Vérifier l'ajustement en utilisant le test de Kolmogorov Smirnov avec $\alpha=5\%$;
3. Calculer la pluie centennale et millénaire en utilisant la loi log normale ;

1. L'ajustement à la loi log normale :

- Classement de l'échantillon par ordre décroissant
- Calcul de la fréquence expérimentale F_i
- Calcul la moyenne et l'écart type
- Tracé de la droite d'ajustement : $\ln(x) = \sigma_{\ln(x)}u + \overline{\ln(x)}$

Tableau 1.9: ajustement de la série d'observations à la loi log normale

Année	x	i	x classée	f(x)	ln(x) classée
1915	526,2	1	1038,7	0,024	6,95
1916	682,5	2	787,3	0,071	6,67
1917	593,6	3	746	0,119	6,61
1918	746	4	682,5	0,167	6,53
1919	318,8	5	593,6	0,214	6,39
1920	524,3	6	526,2	0,262	6,27
1921	389,8	7	524,3	0,31	6,26
1922	486,3	8	513,7	0,357	6,24
1923	396,5	9	504,4	0,405	6,22
1924	504,4	10	502,1	0,452	6,22
1925	299	11	486,3	0,5	6,19
1926	296,1	12	396,5	0,548	5,98
1927	787,3	13	389,8	0,595	5,97
1928	502,1	14	367,4	0,643	5,91
1929	513,7	15	363,9	0,69	5,90
1930	367,4	16	334,9	0,738	5,81
1931	363,9	17	318,8	0,786	5,76
1932	301,4	18	301,4	0,833	5,71
1933	286,9	19	299	0,881	5,70
1934	334,9	20	296,1	0,929	5,69
1935	1038,7	21	286,9	0,976	5,66

La moyenne : $\overline{\ln x} = \Sigma \frac{\ln(x)}{N} = 6,13$

L'écart type : $\sigma_{\ln x} = \sqrt{\frac{1}{N-1} \Sigma_{i=1}^N (\ln x)^2 - N \overline{\ln(x)}^2} = 0,36$

La droite d'ajustement est donnée par l'équation suivante :

$$\ln(x) = 0,36u + 6,13$$

2. Vérification de l'ajustement en utilisant le test de Kolmogorov Smirnov :

Les étapes d'application de ce test sont répertoriées dans le tableau 1.10 :

Tableau 1.10: application du test de Kolmogorov Smirnov

Année	x	i	x classée	ln(x)classée	f (xi)	Ui	f ₀ (xi)	Di+	Di-
1915	526,2	1	286,9	5,659	0,05	-1,28	0,1003	0,0530	0,1003
1916	682,5	2	296,1	5,691	0,10	-1,20	0,117	0,0220	0,0694
1917	593,6	3	299	5,7	0,14	-1,17	0,123	0,0200	0,0278
1918	746	4	301,4	5,708	0,19	-1,15	0,1271	0,0630	0,0158
1919	318,8	5	318,8	5,765	0,24	-0,99	0,1611	0,0770	0,0294
1920	524,3	6	334,9	5,814	0,29	-0,86	0,1977	0,0880	0,0404
1921	389,8	7	363,9	5,897	0,33	-0,63	0,2676	0,0660	0,0181
1922	486,3	8	367,4	5,906	0,38	-0,60	0,2743	0,1070	0,0590
1923	396,5	9	389,8	5,966	0,43	-0,44	0,3336	0,0950	0,0474
1924	504,4	10	396,5	5,983	0,48	-0,39	0,3483	0,1280	0,0803
1925	299	11	486,3	6,187	0,52	0,17	0,5675	0,0440	0,0913
1926	296,1	12	502,1	6,219	0,57	0,26	0,5987	0,0270	0,0749
1927	787,3	13	504,4	6,223	0,62	0,27	0,6064	0,0130	0,0350
1928	502,1	14	513,7	6,242	0,67	0,32	0,6255	0,0410	0,0065
1929	513,7	15	524,3	6,262	0,71	0,38	0,6443	0,0700	0,0224
1930	367,4	16	526,2	6,266	0,76	0,39	0,648	0,1139	0,0663
1931	363,9	17	593,6	6,386	0,81	0,72	0,7704	0,0390	0,0085
1932	301,4	18	682,5	6,526	0,86	1,10	0,8643	0,0070	0,0548
1933	286,9	19	746	6,615	0,91	1,35	0,9099	0,0050	0,0528
1934	334,9	20	787,3	6,669	0,95	1,50	0,9319	0,0200	0,0271
1935	1038,7	21	1038,7	6,946	1,00	2,26	0,9878	0,0120	0,0354
						Dmax		0,128	

La valeur maximale :

$$D_{\max} = 0,128$$

Signification du D_{\max} :

$$\text{Pour } \alpha = 0,05 : D\alpha = \frac{0,895}{s} \text{ avec } s = \sqrt{N} - 0,01 + \frac{0,85}{\sqrt{N}}$$

$$s = \sqrt{21} - 0,01 + \frac{0,85}{21} = 4,61$$

On déduit que :

$$D_{\max} < D_{\alpha} \Rightarrow 0,12 < 0,19 \Rightarrow \text{Ajustement satisfaisant}$$

3. Calcul de la pluie centennale et millénaire pour loi log normale :

La droite d'ajustement : $\ln(x) = 0,36u + 6,13$

La pluie centennale :

$$T = 100 \text{ ans} \longrightarrow FD = 0,01 \longrightarrow FND = 1 - 0,01 = 0,99$$

$$U = 2,33$$

$$\ln x_{0,01} = 6,98 \longrightarrow X_{0,01} = 1062,94 \text{ mm}$$

La pluie millénaire :

$$T = 1000 \text{ ans} \longrightarrow FD = 0,001 \longrightarrow FND = 1 - 0,001 = 0,999$$

$$U = 3,1$$

$$\ln x_{0,001} = 7,24 \longrightarrow X_{0,001} = 1397,44 \text{ mm}$$

Exercice 3

En utilisant les données de l'exercice n°1 :

1. Ajuster la série d'observations à la loi de Gumbel ;
2. Calculer la pluie décennale en utilisant la loi de Gumbel.

Solution

1. Ajustement à la loi de Gumbel

Les étapes d'application de l'ajustement à la loi de Gumbel sont :

- Classement de l'échantillon par ordre croissant
- Calcul des fréquences expérimentales F_i
- Calcul de la variable de Gumbel $y_i = -\ln[-\ln(F_i)]$
- Calcul des paramètres de la loi de Gumbel x_0 et $1/\alpha$

Tableau 1.11 : ajustement des données à la loi de Gumbel

i	P_{ann} (mm)	x classée	$f_i(x)$	y_i	$x_i \cdot y_i$	y_i^2
1	351	319,2	0,02	-1,32	-420,85	1,74
2	469,6	338,6	0,07	-0,97	-328,58	0,94
3	400,2	339,3	0,12	-0,76	-256,27	0,57
4	387,6	351	0,17	-0,58	-204,70	0,34

5	403,7	355,2	0,21	-0,43	-153,47	0,19
6	559,6	360,1	0,26	-0,29	-105,33	0,09
7	355,2	382,9	0,31	-0,16	-61,01	0,03
8	485,1	387,6	0,36	-0,03	-11,31	0,00
9	360,1	400,2	0,40	0,10	40,19	0,01
10	339,3	403,7	0,45	0,23	93,51	0,05
11	423	423	0,50	0,37	155,03	0,13
12	382,9	466,9	0,55	0,51	236,81	0,26
13	634,4	469,6	0,60	0,66	308,17	0,43
14	512,3	485,1	0,64	0,82	396,24	0,67
15	655	491,5	0,69	0,99	488,18	0,99
16	491,5	512,3	0,74	1,19	610,55	1,42
17	535,6	535,6	0,79	1,42	761,78	2,02
18	319,2	559,6	0,83	1,70	952,43	2,90
19	589,7	589,7	0,88	2,07	1218,04	4,27
20	338,6	634,4	0,93	2,60	1650,86	6,77
21	466,9	655	0,98	3,73	2440,30	13,88

La moyenne de x : $\bar{x} = 450,5$

La moyenne de Y : $\bar{y} = 0,56$

$$\sum_{i=1}^N y_i x_i = 7810,559$$

$$\sum_{i=1}^N y_i^2 = 37,687$$

On calcule alors les paramètres de la loi de Gumbel :

$$\frac{1}{\alpha} = r \frac{\sigma_x}{\sigma_y} = \frac{\sum_{i=1}^N y_i x_i - N \bar{x} \bar{y}}{\sum_{i=1}^N y_i^2 - N \bar{y}^2} = 79.84$$

$$x_0 = \bar{y} - \frac{1}{\alpha} \bar{x} = 450.47$$

La droite d'ajustement est donnée par l'équation suivante:

$$x = 79,84y + 450,47$$

2. Calcul de la pluie décennale

$$T = 10 \text{ans} \longrightarrow FD = 0,1 \longrightarrow FND = 1 - 0,1 = 0,9$$

Notions de probabilité et d'analyse statistique appliquée à l'hydrologie

On a : $F(x) = e^{(-e)^{-y}}$

Alors : $Y_{0,1} = -\ln[-\ln(F(x))] \implies Y_{0,1} = -\ln[-\ln(0,9)] = 2,25$

En remplaçant $Y_{0,1}$ par sa valeur dans l'équation d'ajustement :

$X_{0,1} = 630,11 \text{ mm}$

Chapitre 2 : Corrélation et analyse des données

Objectif du chapitre

Les corrélations et l'analyse des données sont d'un grand intérêt en hydrologie pour étudier les relations entre les variables hydrologiques, climatologiques, morphologiques, etc.

Leur utilisation peut se classer sous trois rubriques:

- a) extension dans le temps de séries d'observations hydrologiques qui sont de trop courte durée ou comportent des lacunes;
- b) prévisions de données hydrologiques (apports mensuels, crues, étiages, etc.) en fonction des conditions hydro-météorologiques observées au moment de la prévision;
- c) extension géographique à des bassins non observés des caractéristiques hydrologiques déterminées sur divers bassins versants de régime analogue.

La théorie complète de l'analyse des données est d'un abord assez difficile et il existe déjà à son sujet une abondante littérature. Nous nous bornerons ici à en donner les grandes lignes et insisterons plus particulièrement sur les points qui sont essentiels pour une application correcte de la théorie.

1. Généralités :

Avant d'aborder le problème de l'homogénéisation des données pluviométrique, quelques notions de statistique doivent être rappelées.

1.1 Association entre variable

Soit deux caractères hydrologique X et Y ; dont X est la pluviométrie annuelle à une station A et Y est la pluviométrie annuelle à une station B.

On cherche à mettre en évidence les relations qui existent entre les deux séries d'observations considérées simultanément.

Les observations relatives à ces deux caractères sont représentées sous la forme d'une série statistique double.

On représente graphiquement les séries statistiques doubles en construisant les diagrammes de dispersion.

Après le report des N points, les situations suivantes peuvent se présenter :

1. Les points se dispersent au hasard dans le plan Oxy
2. Les points suivants une ligne très nette
3. Le nuage des points a une forme allongée

En fait, après avoir décelé graphiquement l'existence d'une corrélation linéaire entre deux caractères, il faut déterminer les différents paramètres de la fonction.

1.2 Les données aberrantes

Dans les séries de données hydrologiques se glissent des erreurs qui peuvent avoir des origines multiples (erreur de lecture, erreur de report, détarage de l'appareillage, etc....).

Classiquement nous différencierons:

- Les erreurs accidentelles qui affectent de façon uniforme certaines mesures réparties aléatoirement dans le temps et dans l'espace. Ce type d'erreurs est dû à des raisons variées, généralement inconnues affectant différemment chaque mesure individuelle.
- Les erreurs systématiques qui affectent d'une façon uniforme certaines portions des séries de mesure, mais portions qui sont aléatoirement distribuées dans le temps et l'espace. Ce type d'erreurs est dû généralement au:
 - ⊕ Déplacement du pluviomètre pendant la période d'observation;
 - ⊕ Changement d'environnement du pluviomètre pendant la période d'observation: croissance d'un arbre qui devient obstacle;
 - ⊕ Substitution pendant la période d'observations de l'éprouvette de lecture du pluviomètre par une autre inadéquate.

Au cours de la manipulation des données, certaines valeurs apparaissent comme étant des valeurs singulières par rapport au reste de la série chronologique. Le phénomène des valeurs aberrantes peut entraîner l'hydrologue à des interprétations erronées.

Contrairement aux valeurs erronées, les données exceptionnelles devraient se répéter dans plusieurs stations. Donc, il est indispensable de comparer tous les relevés des stations faisant partie de la même micro- région pour s'assurer qu'ils ne présentent pas de discordance anormale.

2. Généralités sur la corrélation

2.1 Notations et définitions

On considérera deux séries d'observations X et Y , notées :

$$x_1, x_2, \dots, x_n$$

$$y_1, y_2, \dots, y_n$$

Ou avec une notation plus condensée :

$$(x_i)_{i=1,2,\dots,n}$$

$$(y_i)_{i=1,2,\dots,n}$$

On appelle coefficient de corrélation, la quantité r_{xy} définie par :

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} = \frac{Cov(x, y)}{\sigma_x^2 \sigma_y^2} \quad 2.1$$

\bar{x} et \bar{y} : sont les moyennes de x et y

σ_x^2 et σ_y^2 : sont les variances de x et y

$Cov(x, y)$: est la covariance

$$Cov(x, y) = \frac{1}{N} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) \quad 2.2$$

2.2 Analyse primaire de qualité :

r_{xy} est un indice, compris entre -1 et +1, mesurant l'intensité et le sens d'une éventuelle liaison linéaire entre la série des x_i et celle des y_i .

- Si $r_{xy} = +1$, il existe une relation linéaire exacte croissante entre les deux séries : $y_i = ax_i + b, a > 0$.
- Si $r_{xy} = -1$, il existe une relation linéaire exacte décroissante entre les deux séries : $y_i = ax_i + b, a < 0$.
- Si $r_{xy} = 0$, il n'existe pas de relation linéaire
- Si $0 < |r_{xy}| < 1$, cas intermédiaire c'est-à-dire une relation monotone plus ou moins floue selon le signe de r_{xy} .

Ces différents cas sont résumés dans la figure 2.1 :

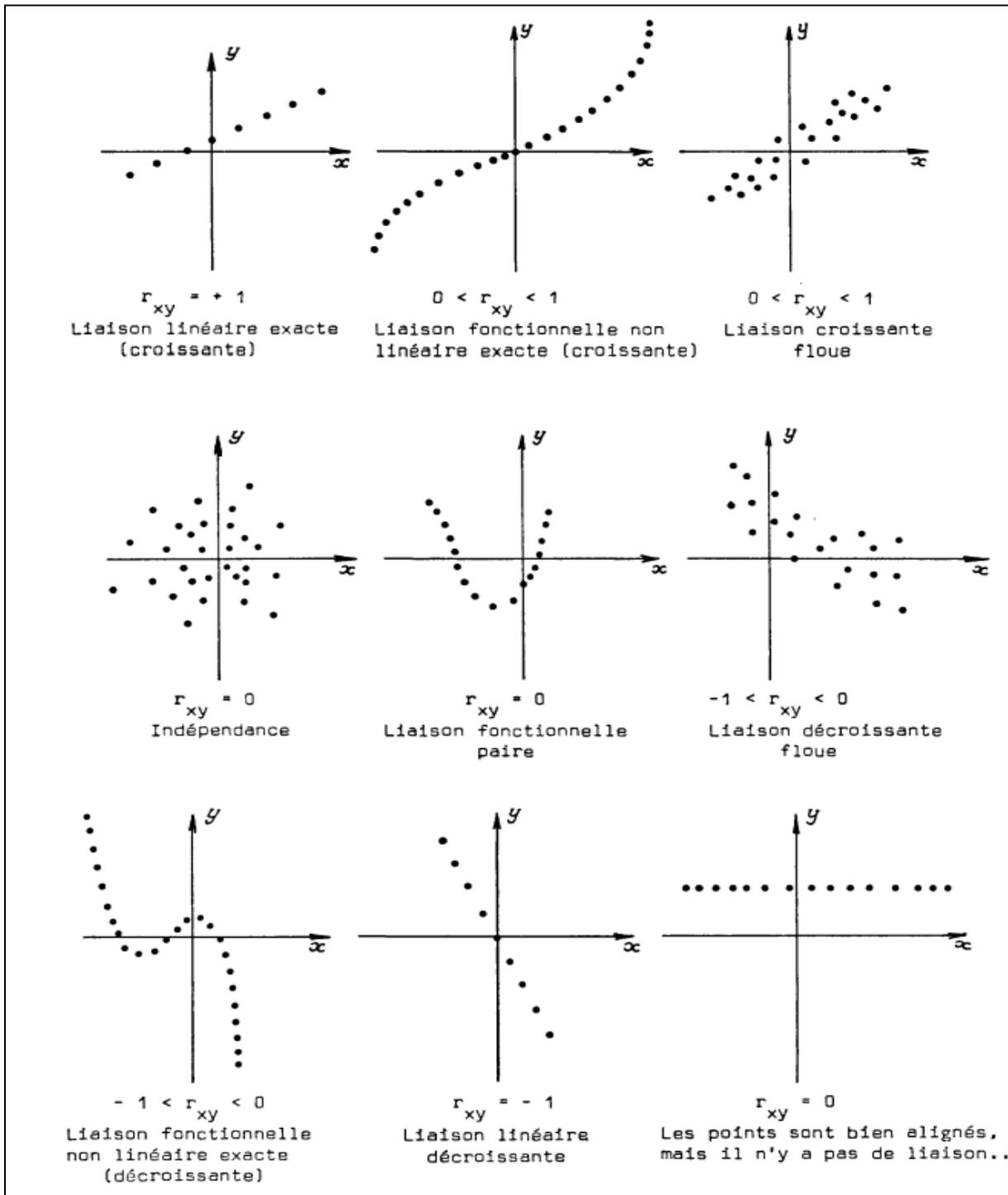


Figure 2.1 : différents cas de la corrélation entre deux variables (Canceill, 1974)

2.3 Complément statistique des données hydrologiques (Renard, 2014)

Plaçons-nous dans la situation suivante : sur un échantillon, nous avons mesuré deux variables quantitatives, que nous noterons X et Y . Nous obtenons donc deux séries de valeurs $x=(x_1, \dots, x_n)$ et $y=(y_1, \dots, y_n)$. Les deux séries peuvent bien sûr être décrites indépendamment

Corrélations et analyse des données

l'une de l'autre, mais nous allons ici nous intéresser à la liaison qui peut exister entre ces variables.

Exemple : Sur un échantillon de 10 petits cours d'eau, on a calculé la superficie du bassin versant (X) ainsi que la crue décennale (Y) :

Tableau 2.1 : Tableau de données

BV	Superficie (km ²)	Q10 (m ³ /s)
1	16.2	5.77
2	21.4	5.37
3	38.7	7.29
4	42.3	12.66
5	55.8	11.09
6	62.1	25.31
7	64	20.18
8	68	39.27
9	73	39.59
10	94	28.30

La manière la plus simple de représenter ces données est de projeter les couples (xi,yi) dans le plan : on obtient ainsi un nuage de points.

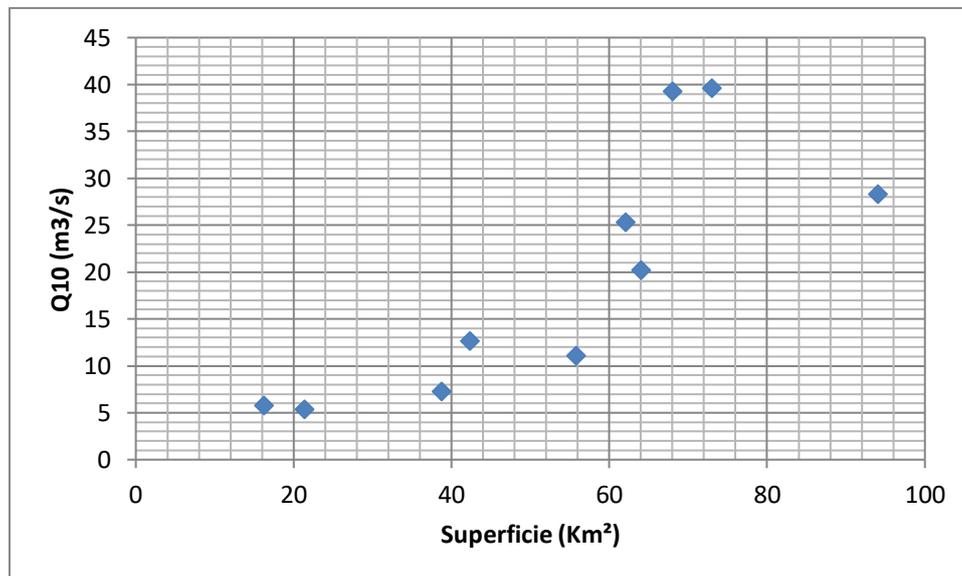


Figure 2.2 : représentation graphique des couples (x, y)

A la vue de ce graphique, il apparaît que les deux variables ont tendance à évoluer conjointement : les « grands » bassins versants ont des crues décennales plus importantes. C'est cette co-évolution que nous allons tenter d'étudier.

Commençons par définir le coefficient de corrélation r_{xy} entre x et y :

$$r_{xy} = \frac{Cov(x,y)}{\sigma_x \sigma_y} \quad 2.3$$

Corrélations et analyse des données

Les valeurs suivantes sont obtenues: $\sigma_x^2 = 24$; $\sigma_y^2 = 13.17$; $Cov(x, y) = 255.01$; $r_{xy} = 0.81$

Ces chiffres confirment la liaison entre les variables. Nous pouvons essayer d'aller plus loin, en cherchant la droite la plus pertinente qui s'ajusterait au nuage de point, c'est en dire en évaluant une relation du type $Y=aX+b$. Evidemment cette relation ne peut pas être parfaite (tous les points ne sont pas alignés), nous introduisons donc des termes d'erreurs, ce qui nous donne la relation :

$$y_i = ax_i + b + E_i; \forall i = 1, \dots, n \quad 2.4$$

Une « bonne » droite permettrait de minimiser ces erreurs. Nous allons donc définir un critère, dit des moindres carrés, construit à partir de la somme des carrés des erreurs :

$$S = \sum_{i=1}^n E_i^2 \quad 2.5$$

Il s'agit donc à présent de trouver les valeurs de a et b minimisant S. Il est aisé de démontrer que ces valeurs sont données par :

$$a = \frac{\sum_{i=1}^N x_i y_i - \bar{x} \bar{y}}{\sum_{i=1}^N x_i^2 - \bar{x}^2} = r_{xy} \frac{\sigma_y^2}{\sigma_x^2} \quad 2.6$$

$$b = \bar{y} - a\bar{x}$$

La qualité de cet ajustement est mesurée par r^2 (qui varie dans $[0,1]$), qui mesure la part de variance expliquée par le modèle linéaire. Voici l'ajustement pour les données précédentes est donné en Figure 2.3.

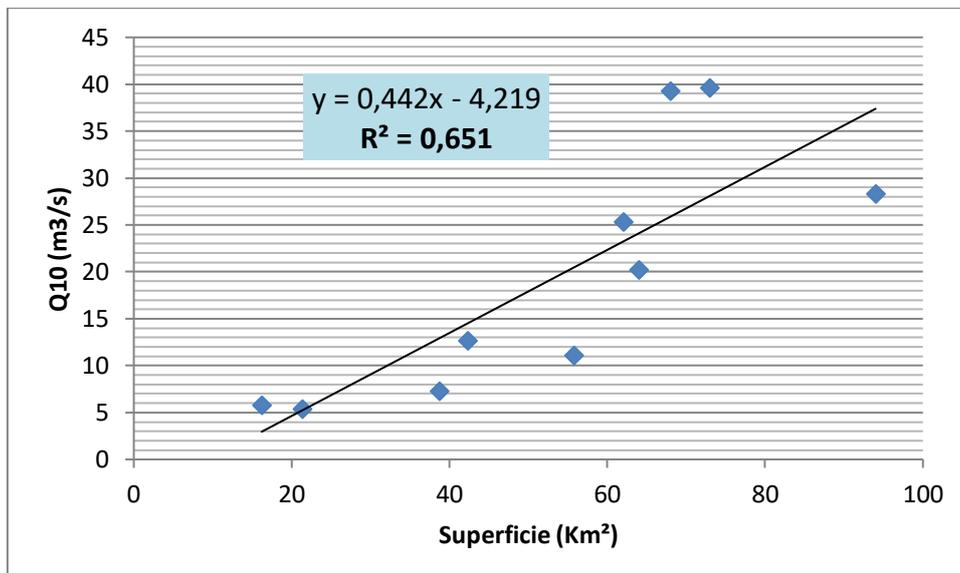


Figure 2.3 : Ajustement de la droite de régression

Une extension importante de la méthode consiste à intégrer plusieurs variables pour en « expliquer » une autre : dans le cas précédent, il pourrait ainsi être bénéfique d'intégrer la pluviométrie ou la nature et l'occupation du sol pour améliorer l'explication du débit décennal. On parle de régression multiple. Ce type d'approche est très important en hydrologie pour les problèmes de régionalisation.

3. Processus, variables et séries

3.1 Définition

Une série chronologique est une succession d'événements pluvieux observés sur un site pendant une durée donnée, ou générés numériquement, pouvant être utilisée en entrée d'un modèle hydrologique en tenant compte des périodes de temps sec qui les séparent.

3.2 Valeurs caractéristiques d'une série chronologique (Grisolet, 1972)

Un examen de la représentation graphique de la série permet de constater la vraisemblance de certaines composantes fondamentales. Ainsi, on peut distinguer :

1. Un mouvement de tendance générale « trend » : la variation tend à se maintenir pendant de nombreuses années dans un sens déterminé.
2. Un mouvement dit « cyclique » qui désigne des oscillations de périodes et d'amplitudes plus ou moins irrégulières.
3. Un mouvement périodique qui se produit à des intervalles de temps égaux avec une forme constante ; la période qui sépare deux points homologues consécutifs peut être la journée (mouvement diurne), le mois, l'année ; dans une période d'une année se trouve le mouvement saisonnier.
4. Un mouvement aléatoire qui échappe à toute prévision et qui entraîne des variations accidentelles qualifiées de variations résiduelles.

Ces diverses composantes : tendance (T), mouvement cyclique (C), variation saisonnière (S), variations résiduelles (R) forment les données brutes $F(t)$; on admet schématiquement qu'elles sont conjuguées ; soit sous la forme additive :

$$F(t)=T(t)+C(t)+S(t)+R(t) \quad 2.7$$

Soit sous la forme multiplicative :

$$F(t)=T(t) \times C(t) \times S(t) \times R(t) \quad 2.8$$

Remarque : les variations dans le temps des phénomènes naturels sont plus complexes que ne l'indiquent l'une ou l'autre de ces formes théoriques.

3.3 Recherche de la tendance générale

La tendance générale traduit l'évolution du phénomène au cours d'une longue période ; sa connaissance peut servir à fixer l'origine à partir de laquelle seront évalués les autres mouvements.

Procédé graphique : si le graphique représentatif des observations fait apparaître une tendance, on trace la courbe qui s'accorde le mieux avec les observations ; il est possible d'y remédier par le procédé des points médians qui consiste à joindre les maximums M_1, M_2, \dots d'une part et les minimums m_1, m_2, \dots d'autre part. Sur les ordonnées qui correspondent à un maximum ou à un minimum, on détermine les points A, B, .. équidistants de la ligne des maximums et de celle des minimums ; la ligne qui joint ces points médians est une ligne de tendance.

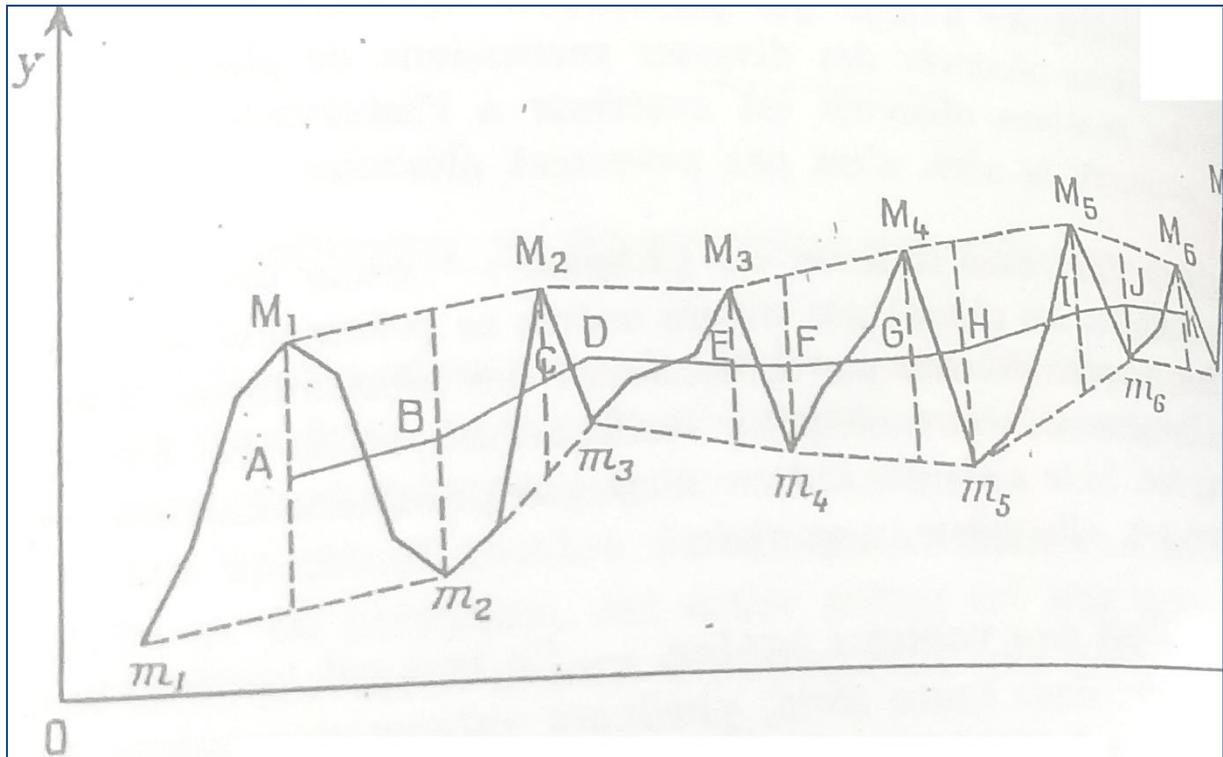


Figure 2.4 : Procédé graphique de recherche de la tendance (Grisolet, 1972)

L'inconvénient du procédé est de donner aux points une importance excessive et de ne pas permettre l'utilisation des données du début et de la fin de la série.

Remarque :

Une méthode plus adéquate encore consiste à déterminer les paramètres de la ligne de tendance au moyen de la méthode des moindres carrés. Ainsi, dans le cas où la tendance paraît rectiligne, l'équation de la droite de tendance sera $y=a+bt$, avec :

$$a = \bar{y} - b\bar{t} \quad \text{et} \quad b = \frac{\sum(y_i - \bar{y})(t_i - \bar{t})}{\sum(t_i - \bar{t})^2} \quad 2.9$$

\bar{y} et \bar{t} étant les moyennes des valeurs de la variable Y et du temps t.

La droite de tendance est la droite de régression par rapport au temps des valeurs observées de la variable.

Pour les calculs, b sera pris de préférence sous la forme :

$$b = \frac{\sum yt - N\bar{y}\bar{t}}{\sum t^2 - N\bar{t}^2} = \frac{\sum yt - \bar{t}\sum y}{\sum t^2 - \bar{t}\sum t} \quad 2.10$$

Dans le cas où l'on fait correspondre l'origine des temps avec la date de la première observation, t prend des valeurs successives 0, 1, 2, ..., N-1 et l'on a :

$$\sum t = \frac{(N-1)N}{2} \quad \text{et} \quad \sum t^2 = \frac{(N-1)N(2N-1)}{6} \quad 2.11$$

D'où :

$$i = \frac{N-1}{2} \quad \text{et} \quad \sum t^2 - \bar{t}\sum t = \frac{N(N^2-1)}{12} \quad 2.12$$

Et, enfin :

$$a = \bar{y} - b\bar{t} = \bar{y} - \frac{N-1}{2}b \quad 2.13$$

Remarque :

Lorsque l'évolution de la grandeur paraît traduire un taux d'accroissement à peu près constant, il convient d'envisager une tendance représentée par une loi de la forme :

$$\log y = a + bt \quad 2.14$$

La méthode de calcul de a et b est la même que précédemment, en remplaçant y par log y.

3.4 Recherche diurne et composante saisonnière

On peut éliminer le mouvement diurne en n'utilisant que les observations qui correspondent à un même instant du cycle, par exemple, ne conserver qu'une observation par jour effectuée à une heure fixe, ou également représenter chaque jour par la valeur moyenne d'observations quotidiennes.

La composante saisonnière représente un mouvement périodique dont la période est égale à une année. Pour les variables hydrologiques, le mouvement saisonnier ne se reproduit pas d'une manière identique chaque année, les extremums n'ayant pas lieu à la même date et avec une amplitude variant d'une année à l'autre. Toutefois, le mouvement conserve une analogie de forme suffisante pour qu'on puisse parler de mouvement saisonnier.

La détermination de la variation saisonnière est faite à partir d'une analyse de la série statistique qui se rapproche à une unité de temps plus courte, le plus souvent on a recours aux moyennes mensuelles.

Partant de N années d'observations, on calcule la moyenne des N valeurs de chacun des 12 mois d'où 12 valeurs moyennes mensuelles $\bar{m}_1, \bar{m}_2, \dots, \bar{m}_{12}$ et une moyenne générale :

$$\bar{M} = \frac{\bar{m}_1 + \dots + \bar{m}_{12}}{12} \quad 2.15$$

La variation saisonnière est pour le mois d'ordre k caractérisée par la différence

$$m_k - \bar{M} \quad 2.16$$

S'il existait une tendance séculaire notable, il serait nécessaire de déterminer d'abord la droite de tendance et de remplacer les données brutes par les données corrigées de la tendance.

3.5 Recherche de la composante cyclique

L'examen graphique établi sur une très longue série d'observations laisse apparaître des oscillations à longues périodes qui se superposent au mouvement d'ensemble. Il est en général difficile et délicat de les analyser quantitativement. Sur une période aussi longue que possible, on détermine la tendance générale (T) et la composante saisonnière (S).

Les données sont ensuite traitées par la méthode des moyennes mobiles pour éliminer l'influence des variations accidentelles. Le rapprochement entre ces valeurs régularisées et la résultante de T et de S fait apparaître la composante cyclique (C).

3.6 Caractère aléatoire des séries d'observations

Le caractère aléatoire simple d'une série d'observations est une hypothèse fondamentale pour l'analyse statistique d'une telle série. De nombreuses méthodes statistiques utilisées soit pour vérifier des propriétés particulières des séries d'observations, soit pour estimer certains paramètres caractéristiques du processus qui produit ces séries ne sont applicables qu'à condition que cette première hypothèse soit effectivement réalisée. Sa vérification constitue donc le problème à résoudre avant de procéder à toute analyse statistique complémentaire.

Les tests les plus répandus portent sur la constance de la moyenne de la série tout au long de sa période d'observations. Ces tests sont en général assez puissants pour faire une distinction entre le caractère aléatoire et le caractère non aléatoire de la série. En revanche, tous ne permettent pas d'identifier une alternative à la constance du type tendance, discontinuité brutale, oscillations ...

Quelques tests ont pour objet la constance de la dispersion de la série, c'est-à-dire qu'ils étudient si la variabilité de la série est uniforme dans le temps.

Soit la série chronologique (x_i) , $i = 1, N$, les x_i désignent les réalisations de la variable X observées à des pas de temps successifs égaux (**Kotz et al., 1981**).

L'hypothèse nulle est donc : « La série des (x_i) , $i=1, N$, est aléatoire ».

3.6.1 Test du rapport de Von Neumann (Buishand, 1982)

Il s'agit du rapport de la moyenne du carré des différences successives des valeurs observées à la variance.

On note V ce rapport :

$$V = \frac{N}{N-1} \cdot \frac{\sum_{i=1}^{N-1} (x_i - x_{i+1})^2}{\sum_{i=1}^N x_i^2 - \frac{1}{N} \left(\sum_{i=1}^N x_i \right)^2} \quad 2.17$$

Pour N grand (N>30), si la série est aléatoire, V est distribué selon une loi normale de moyenne $\frac{2N}{N-1}$ et de variance approximativement égale à $\frac{4(N-2)}{(N-1)^2}$.

Il en résulte que, si l'hypothèse nulle est vraie, la variable :

$$U = \frac{V - \frac{2N}{N-1}}{\frac{2\sqrt{N-2}}{N-1}} \quad 2.18$$

Pour un risque α de première espèce donné, la région d'acceptation de l'hypothèse nulle est comprise entre :

$$(V)_t^- = \frac{2N - 2U_{1-\alpha/2}\sqrt{N-2}}{N-1} \quad \text{et} \quad (V)_t^+ = \frac{2N + 2U_{1-\alpha/2}\sqrt{N-2}}{N-1} \quad 2.19$$

aucune hypothèse alternative spécifique n'est associée à ce test.

3.6.2 Autocorrélogramme (Lubes et al, 1994)

On suppose qu'il existe des dépendances significatives entre les termes successifs d'une série non aléatoire.

Le coefficient d'autocorrélation d'ordre k est donné par l'expression :

$$r_k = \frac{\sum_{t=1}^{N-k} (x_t - \bar{x}_1) \cdot (x_{t+k} - \bar{x}_2)}{\sqrt{\left[\sum_{t=1}^{N-k} (x_t - \bar{x}_1)^2 \sum_{t=1}^{N-k} (x_{t+k} - \bar{x}_2)^2 \right]}} \quad 2.20$$

Avec \bar{x}_1 moyenne des observations (x_i) , $i = 1, N-k$, et \bar{x}_2 moyenne des observations (x_i) , $i = k+1, N$.

D'après **Chatfield** (1989), si une série chronologique est aléatoire, alors pour N grand, $r_k \approx 0$ pour toute valeur de k non nulle. En fait pour une série chronologique aléatoire, et pour N grand, r_k suit approximativement une distribution normale de moyenne nulle et de variance $1/N$.

Il est donc possible de définir une région de confiance contenant pour un seuil de confiance donné, sous l'hypothèse nulle, l'autocorrélogramme. Pour un seuil de confiance $1-\alpha/2$ donné, la région de confiance est définie par $\pm \frac{U_{1-\alpha/2}}{\sqrt{N}}$. U désigne la variable normale réduite.

Une importance particulière doit être donnée au comportement de l'autocorrélogramme pour de faibles valeurs de k , notamment pour $k=1$ (**WMO, 1966**). En effet sur les vingt premières valeur de r_k , il n'est pas rare qu'une valeur sorte de la région de confiance même lorsque la série est réellement aléatoire. Ceci souligne les difficultés d'interprétation de l'autocorrélogramme. En revanche l'estimation de ce dernier est incontournable comme première exploitation de toute série chronologique.

3.6.3 Test de corrélation sur le rang

Ce test se propose de calculer le nombre de paires P pour lesquelles $x_j > x_i$, $j > i$, avec $i = 1, \dots, N-1$.

Pour N grand, sous l'hypothèse nulle, la variable τ telle que :

$$\tau = 1 - \frac{4Q}{N(N-1)} \quad \text{avec} \quad Q = \frac{N(N-1)}{2} - P \quad 2.21$$

suit une distribution normale de moyenne nulle et de variance égale à $\sigma_r^2 = \frac{2(2N+5)}{9N(N-1)}$

Il en résulte que si l'hypothèse nulle est vraie, la variable $U = \tau / \sigma_r$ est une variable normale réduite.

Pour un risque α de première espèce donné, la région d'acceptation de l'hypothèse nulle est comprise entre :

$$-U_{1-\alpha/2\sigma_r} \quad \text{et} \quad +U_{1-\alpha/2\sigma_r}$$

L'hypothèse alternative reconnue de ce test est celle d'une tendance.

Lorsque l'on s'intéresse directement à la distribution asymptotique de la variable P , ce test porte le nom de test de **Mann-Kendall**. Les deux formulations sont équivalentes.

3.6.4 Statistique de rang de Spearman (Sneyers S, 1975)

Les observations originales (x_i) , $i = 1, N$ sont remplacées par le rang y_i qui leur est attribué lorsqu'on les range par ordre de grandeur croissante et la statistique du test est le coefficient de corrélation r_s entre les séries i et y_i , coefficient qui peut se calculer au moyen de la formule :

$$r_s = 1 - \frac{6}{n(n^2-1)} \sum (y_i - i)^2 \quad 2.22$$

La distribution de cette statistique est asymptotiquement normale avec $E(r_s)=0$ et $\text{var}(r_s) = \frac{1}{n-1}$.

Il convient de déterminer à l'aide d'une table de la loi normale réduite la probabilité α_1 tel que

$$\alpha_1 = P\left(|u| \geq |u(r_s)|\right) \quad \text{avec } u(r_s) = r_s \sqrt{n-1} \quad 2.23$$

L'hypothèse nulle est acceptée ou rejetée au niveau $\alpha_1 > \alpha_0$ ou $\alpha_1 < \alpha_0$.

Ce test est largement cité dans la littérature. L'hypothèse alternative est celle de la tendance.

La puissance de ce test est comparable à celle du test de corrélation sur le rang.

3.7 Tests de détection de ruptures

« **Rupture** » doit être compris comme un changement dans la loi de probabilité de la série chronologique à un instant donné, le plus souvent inconnu (**Lubes et al., 1994**).

3.7.1 Test de Pettit « version modifiée de Mann-Whitney » (**Lubes et al., 1994**)

Le test de Pettitt est non-paramétrique et dérive du test de **Mann-Whitney (1947)**. L'absence d'une rupture dans la série (x_i) de taille N constitue l'hypothèse nulle. Le fondement de ce test est le suivant :

La série est divisée en deux sous-échantillons respectivement de taille m et n .

Les valeurs des deux échantillons sont regroupées et classées par ordre croissant. On calcule alors la somme des rangs des éléments de chaque sous-échantillon dans l'échantillon total. Une statistique est définie à partir des deux sommes ainsi déterminées, et testée sous l'hypothèse nulle d'appartenance des deux sous-échantillons à la même population.

La formulation du test de Mann-Whitney modifiée par Pettit est la suivante :

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

$$\text{Soit } D_{ij} = \text{sgn}(x_i - x_j) \text{ avec : } \begin{cases} \text{sgn}(x) = 1 & \text{si } x > 0 \\ \text{sgn}(x) = 0 & \text{si } x = 0 \\ \text{sgn}(x) = -1 & \text{si } x < 0 \end{cases} \quad 2.24$$

On considère $U_{t,N}$ telle que :

$$U_{t,N} = \sum_{i=1}^t \sum_{j=t+1}^N D_{ij} \quad 2.25$$

Soit K_N la variable définie par le maximum en valeur absolue de $U_{t,N}$ pour t variant de 1 à N .

A partir de la théorie des rangs, Pettitt montre que 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 :

$$\text{prob}(K_N > k) \approx 2 \exp\left(\frac{-6k^2}{(N^3 + N^2)}\right) \quad 2.26$$

Pour un risque α de première espèce donné, si $\text{prob}(K_N > k)$ est inférieur à α , l'hypothèse nulle est rejetée. La série comporte alors une rupture localisée au moment τ où est observée K_N .

Le test est plus particulièrement sensible à un changement de moyenne.

3.7.2 Test du rapport de vraisemblance (Buishand, 1982, 1984)

Ce test fait référence au modèle simple qui suppose un changement de moyenne de la série :

$$x_i = \begin{cases} \mu + \varepsilon_i & i=1, \dots, m \\ \mu + \Delta + \varepsilon_i & i=m+1, \dots, n \end{cases} \quad 2.27$$

Les ε_i sont de variables aléatoires normales de moyenne nulle et de variance commune inconnue σ^2 . Le point de rupture m et les paramètres μ et Δ sont aussi inconnus.

Plusieurs méthodes statistiques ont été développées pour tester l'hypothèse nulle $\Delta = 0$ contre l'hypothèse alternative $\Delta \neq 0$.

On s'intéresse aux termes du cumul d'écart suivants :

$$S_k^* = \sum_{i=1}^k (x_i - \bar{x}) \text{ Pour } k=1, \dots, N \quad 2.28$$

$$S_0^* = 0$$

\bar{x} est la moyenne des valeurs x_1, x_2, \dots, x_N .

S_k^* est tel que :

$$E(S_k^*) = -k(N-m)N^{-1}\Delta, \quad k=0, \dots, m \quad 2.29$$

$$E(S_k^*) = -m(N-k)N^{-1}\Delta, \quad k=m+1, \dots, N \quad 2.30$$

$$\text{var}(S_k^*) = k(N-k)N^{-1}\sigma^2, \quad k=0, \dots, N \quad 2.31$$

On observe que la moyenne des S_k^* est :

- Nulle pour une série homogène ($\Delta=0$) ;
- Positive pour $\Delta < 0$;
- Négative pour $\Delta > 0$.

La variance est maximale si $k=N/2$. Même pour une série purement aléatoire, les valeurs de S_k^* peuvent différer de zéro, spécialement pour k au voisinage de $N/2$.

Le test du rapport de vraisemblance porte sur la variable :

$$V = \max \left\{ \frac{|S_k^*|}{D_x \sqrt{k(N-k)}} \right\} \text{ Pour } 1 \leq k \leq N-1 \quad 2.32$$

avec D_x écart-type de la série.

De grandes valeurs de V conduisent à rejeter l'hypothèse nulle. Les valeurs critiques de la statistiques V peuvent être obtenues à partir des valeurs critiques de la statistique W telle que :

$$W = \frac{\sqrt{N-2} \cdot V}{\sqrt{1-V^2}} \quad 2.33$$

qui ont été tabulées par Worsley (1979).

La statistique V est fortement dépendante de l'hypothèse de normalité sur la distribution de la variable étudiée.

3.7.3 Statistique U (Buishand, 1982, 1984)

Le test ici présent est de nature Bayésienne. Il fait référence au même modèle de base et aux mêmes termes que le test précédent.

En supposant une distribution a priori uniforme pour la position du point de rupture m , la statistique U est définie par :

$$U = \frac{1}{N(N+1)} \sum_{k=1}^{N-1} \left(\frac{S_k^*}{D_x} \right)^2 \quad 2.34$$

Des valeurs critiques de la statistique U sont données par Buishand (1982) à partir d'une méthode de Monte Carlo. De meilleures estimations sont parues ultérieurement (Buishand, 1984) (tableau n°III.3).

Tableau 2.2: Valeurs critiques et bornes de la statistique U pour tester $\Delta = 0$ contre $\Delta \neq 0$.

Taille de l'échantillon	Borne inférieure Umin	Niveau de signification α			Borne supérieure Umax
		0.10	0.05	0.01	
10	0.023	0.333	0.416	0.574	0.929
20		0.340	0.440	0.659	
30	0.012	0.343	0.447	0.688	1.934
40		0.344	0.451	0.702	
50	0.008	0.345	0.453	0.710	2.944
100		0.346	0.457	0.727	
∞	0.006	0.347	0.461	0.743	3.956
	0.005				4.968
	0.002				10.033
	0				∞

La statistique U donne moins de poids que la statistique V (test précédent) aux premières et dernières valeurs de la série. En conséquence, la statistique V est supérieure à la statistique U pour déceler un changement de moyenne en début et en fin de série. Pour tout changement de moyenne survenant au milieu de la série, la statistique U s'avère plus performante.

En cas de rejet de l'hypothèse nulle, aucune estimation de la date de rupture n'est proposée par ce test.

Outre cette procédure, la construction d'une ellipse de contrôle permet d'analyser l'homogénéité de la série de (x_i) . la variable S_k , définie au dessus, suit une distribution normale de moyenne nulle et de variance $k(N-k)N^{-1}\sigma^2$, $k = 0, \dots, N$ sous l'hypothèse nulle d'homogénéité de la série des (x_i) . il est donc possible de définir une région de confiance dite ellipse de contrôle associée à un seuil de confiance contenant la série des S_k .

3.7.4 La Procédure Bayésienne de Lee et Heghinian (Lubes et al., 1994)

Le modèle de base de la procédure est le suivant :

$$x_i = \begin{cases} \mu + \varepsilon_i & i=1, \dots, \tau \\ \mu + \delta + \varepsilon_i & i=\tau+1, \dots, n \end{cases} \quad 2.34$$

Les ε_i sont des variables aléatoires normales de moyenne nulle et de variance σ^2 . τ, μ, δ et σ sont des paramètres inconnus et indépendants, $1 \leq \tau \leq N-1$, $-\infty \leq \mu \leq +\infty$, $-\infty \leq \delta \leq +\infty$, $\sigma > 0$.

τ et δ représentent respectivement la position dans le temps et l'amplitude d'un changement éventuel de moyenne.

L'approche Bayésienne présentée ici est fondée sur les distributions marginales *a posteriori* de τ et δ (Lee et Heghinian, 1977).

a. Les distributions *a priori* de τ et δ sont :

$$p(\tau) = \frac{1}{N-1}, \quad \tau=1, 2, \dots, N-1 \quad 2.35$$

$p(\tau)$ est normale de moyenne nulle et de variance σ_δ^2 .

b. La distribution *a posteriori* de τ est définie par :

$$p(\tau/x) = [N/(\tau(N-\tau))]^{1/2} [R(\tau)]^{-(N-2)/2}, \quad 0 \leq \tau \leq N-1 \quad 2.36$$

avec :

$$R(\tau) = \frac{H(\tau)}{\sum_{i=1}^N (x_i - x_N)^2} = \frac{\sum_{i=1}^{\tau} (x_i - \bar{x}_\tau)^2 + \sum_{i=\tau+1}^N (x_i - \bar{x}_{N-\tau})^2}{\sum_{i=1}^N (x_i - \bar{x}_N)^2} \quad 2.37$$

et :

$$\bar{x}_N = \frac{1}{N} \sum_{i=1}^N x_i ; \quad \bar{x}_\tau = \frac{1}{\tau} \sum_{i=1}^{\tau} x_i ; \quad \bar{x}_{N-\tau} = \frac{1}{N-\tau} \sum_{i=\tau+1}^N x_i \quad 2.38$$

c. la distribution *a posteriori* de δ est définie à partir de la distribution *a posteriori* de τ , $p(\tau/x)$ et de la distribution conditionnelle *a posteriori* par rapport à τ ,

$p(\delta/\tau, x)$:

$$p(\delta/x) = \sum_{\tau=1}^{N-1} p(\delta/\tau, x) \cdot p(\tau/x) \quad 2.39$$

d. la distribution conditionnelle *a posteriori* de δ par rapport à τ est une distribution de Student de moyenne $\hat{\delta}_\tau = \bar{x}_N - \bar{x}_\tau$ et de variance $\sigma_{\delta/\tau}^2 = \frac{N.H(\tau)}{(N-2) \cdot (\tau(N-\tau))}$ avec $\nu = N-2$, la fonction densité de probabilité de cette loi de Student est la suivante :

$$p(\delta/\tau, x) = \frac{\nu^{\nu/2} \Gamma((\nu+1)/2)}{\Gamma(1/2) \cdot \Gamma(\nu/2) \cdot (\sigma_{\delta/\tau}^2)^{1/2}} \cdot \frac{1}{(\nu + (\delta - \hat{\delta}_\tau)^2 / \sigma_{\delta/\tau}^2)^{(\nu+1)/2}} \quad 2.40$$

Le changement éventuel, position et amplitude, correspond au mode des distributions *a posteriori* de τ et δ . La méthode fournit la probabilité que le changement se produise au moment τ dans une série où on suppose *a priori* qu'il y a effectivement un changement à un moment indéterminé. De même elle donne une estimation de la probabilité que l'amplitude du changement ait la valeur δ .

3.7.5 La procédure de segmentation des séries hydrométéorologiques (Hubert P., 1997)

Une procédure de segmentation de séries hydrométéorologiques a été présentée par **Hubert P** en 1989.

Le principe de cette procédure est de « découper » la série en m segments ($m > 1$) de telle sorte que la moyenne calculée sur tout segment soit significativement différente de la moyenne (ou des) segment(s) voisin(s). Une telle méthode est appropriée à la recherche de multiples changements de moyenne (Hubert et al, 1997 ; 1998).

La segmentation est définie de la façon suivante :

Toute la série x_i , $i = i_1, i_2$ avec $i_1 \leq 1$ et $i_2 \geq N$ où ($i_1 < i_2$) constitue un segment de la série initiale des (x_i) , $i = 1, \dots, N$.

Toute partition de la série initiale en m segments est une segmentation d'ordre m de cette série.

A partir d'une segmentation particulière d'ordre m pratiquée sur la série initiale, on définit :

- i_k , $k=1, 2, \dots, m$, le rang dans la série initiale de l'extrémité terminale du $k^{\text{ième}}$ segment ;
- $n_k = i_k - i_{k-1}$, la longueur du $k^{\text{ième}}$ segment ;

- \bar{x}_k la moyenne du $k^{\text{ième}}$ segment, $\bar{x}_k = \frac{\sum_{i=i_{k-1}+1}^{i=i_k} x_i}{n_k}$;
- D_m , l'écart quadratique entre la série et la segmentation considérée, $D_m = \sum_{k=1}^m d_k$
avec $d_k = \sum_{i=i_{k-1}+1}^{i=i_k} (x_i - \bar{x}_k)^2$. Cet écart permet d'apprécier la proximité de la série et de la segmentation qui lui est appliquée.

La segmentation retenue au terme de la mise en œuvre de la procédure doit être telle que pour un ordre m de segmentation donné, l'écart quadratique D_m soit minimum. Cette condition est nécessaire mais non suffisante pour la détermination de la segmentation optimale. Il faut lui adjoindre la contrainte suivante selon laquelle les moyennes de deux segments contigus doivent être significativement différentes : $\bar{x}_k \neq \bar{x}_{k+1} \forall k=1,2,\dots,m-1$. Cette contrainte est satisfaite par application du test de Scheffé H (1953) qui repose sur le concept de contraste (**Dagnélie P, 1970 ; 1975**).

Par conséquent si lors du processus de segmentation d'ordre $m+1$, aucune segmentation produite n'est valide au sens du test de Scheffé, la segmentation de la série qui est retenue en tant que meilleure segmentation est la segmentation optimale d'ordre m .

D'après les auteurs (**Hubert P et al., 1989**), cette procédure de segmentation peut être regardée comme un test de stationnarité, « la série étudiée est stationnaire », constituant l'hypothèse nulle de ce test. Si la procédure ne produit pas de segmentation acceptable d'ordre supérieur ou égal à 2, l'hypothèse nulle est acceptée.

3.7.6 Les conditions d'application des méthodes

Les procédures statistiques qui sont utilisées regroupent des tests statistiques classiques, une méthode bayésienne et une technique de segmentation de séries chronologiques.

Leurs conditions d'application sont ici précisées.

a. Les tests classiques

Le test de corrélation sur le rang ne suppose aucune propriété particulière de la série chronologique étudiée. Le test de Pettitt s'applique à des séries non autocorrélées et requiert implicitement que la variance de la série ne soit pas affectée par la rupture si une rupture prioritairement recherchée. La statistique de Buishand repose sur l'ensemble des hypothèses suivantes : normalité de la série, égalité de variances des distributions de part et d'autre du point de rupture, absence d'autocorrélation.

b. Procédure bayésienne

La procédure bayésienne de Lee et Heghinian impose normalité, non autocorrélation et constance de la variance.

c. Segmentation

La segmentation qui fait intervenir le test de Scheffé sous-entend implicitement la normalité de la série chronologique.

3.8 Lissage par la méthode de la Moyenne Mobile

La moyenne mobile ou "moyenne glissante" Une Moyenne Mobile (MM) est une courbe calculée à partir des moyennes d'une valeur, sur une période donnée: chaque point d'une moyenne mobile sur 100 séances est la moyenne des 100 derniers cours de la valeur considérée.

La moyenne mobile est calculée par une moyenne arithmétique, calculée sur la valeur de cette observation et les valeurs voisine qui l'encadrent. Dans le cadre de ce cours, on a utilisé 3 valeurs ainsi x_i est remplacé dans le série par :

$$MM = \frac{x_{i-1} + x_i + x_{i+1}}{3} \quad 2.41$$

Où : x_i est la pluviométrie annuelle (P_i).

La moyenne mobile peut être calculée sur différentes périodes, ce qui permet de dégager des tendances à court terme, à moyen terme ou à long terme. Cette courbe, affichée simultanément avec la courbe d'évolution des valeurs, permet de lisser les variations des valeurs, et de dégager une tendance en hausse si la moyenne mobile est croissante (valeur > à MM) et en baisse si la moyenne mobile est décroissante (valeur < MM).

Cette méthode réduit l'influence des variations accidentelles, élimine l'effet des fluctuations de très courtes durées et par conséquent les résultats sont plus commodes à interpréter. Cette méthode, permet aussi, de faire ressortir les grandes tendances pluviométriques dans les séries météorologiques.

4. Test d'homogénéité

L'étude hydrologique d'un bassin ou d'une région est basée sur l'exploitation des données d'observations recueillies sur des périodes plus ou moins longues, continues ou non. Les méthodes statistiques d'analyse de ces séries exigent de celles-ci une homogénéité de leurs composants.

4.1 Définition de la notion d'homogénéité

Le problème d'homogénéité est le suivant: disposant de plusieurs échantillons distincts et indépendants, on veut savoir s'ils sont tirés de la même population. C'est par exemple, le cas où l'on est en présence des observations d'une série chronologique qui ont été prélevées successivement dans une population de nature déterminée, et on se demande si, à un moment donné des prélèvements, la population en question s'est modifiée.

L'hétérogénéité des observations d'une série chronologique peut provenir de plusieurs raisons, entre autres, dans une analyse des caractéristiques d'un phénomène hydrologique:

- les conditions climatiques sont-elles les mêmes pour une période donnée?
- un fait nouveau est-il venu perturber la population dont l'échantillon est extrait ?
- les observations de l'échantillon ont-elles évoluées sous l'influence d'une modification des conditions de leur mesure (déplacement du site, changement d'appareils etc.) ?
- assiste-t-on à un changement du climat?

Dans la formulation statistique du problème d'homogénéité, on considère k populations aléatoires définies par les k fonctions de répartition $F_j(x)$, $j = 1, \dots, k$. On extrait « au hasard » dans chaque population un échantillon de taille n_j .

On dispose ainsi de k échantillons indépendants extraits au hasard, de sorte que l'on cherche à vérifier les hypothèses suivantes:

$$H_0: F_i(x) = \dots = F_j(x) = \dots = F_k(x) \quad 2.42$$

$$H_1: \text{il existe } i \text{ et } j \text{ tel que } F_i(x) \neq F_j(x) \quad 2.43$$

En général, cette étude se limite toujours à deux populations, c'est-à-dire $k = 2$. Dans le système précédent, l'hypothèse nulle H_0 affirme que les k échantillons proviennent d'une même population (les observations de la série chronologique sont homogènes), tandis que l'hypothèse alternative met en doute cette assertion (les observations de la série chronologique ne sont pas homogènes).

4.2 Approche générale

Les méthodes utilisées pour vérifier l'homogénéité des séries hydrologiques peuvent être analytiques ou graphiques.

4.2.1 Tests graphiques

4.2.1.1 Le simple cumul

Le principe consiste à cumuler les valeurs observées et voir ensuite si ces cumuls forment une droite en fonction du temps, avec ou sans changement de pente.

D'après cette méthode, une série est dite homogène si la droite obtenue ne représente pas de cassure apparente, c'est à dire que dans le cas contraire la série est hétérogène.

Les années sont portées en abscisse et les cumuls en ordonnée. L'application de cette méthode est limitée, lorsqu'il s'agit d'une série dont les valeurs varient fortement d'une année à une autre, et même s'il existe une cassure de la droite, on ne peut pas différencier entre une hétérogénéité des données et une tendance climatique.

4.2.1.2 Test du double cumul

Les valeurs correspondantes à la même période d'observations sont reportées en coordonnées rectangulaires, obtenant une courbe de double cumul. Si les données de la station contrôlée sont homogènes par rapport à celles de la station de base, la courbe des doubles cumuls avoisine une droite (Fig. 2.5). Si elle possède une cassure à partir d'un point M, les observations à partir de ce point sont hétérogènes. Dans le cas où l'hétérogénéité est détectée, la correction s'effectue par modification de la pente de la droite de double cumul des données antérieures ou postérieures à la date de la cassure. Seul le but visé par l'étude en cours peut indiquer quelle partie de la série est à corriger. On corrige les données observées en multipliant le rapport de pente $\frac{m_1}{m_2}$ ou $\frac{m_2}{m_1}$ par la valeur erronée dans la série observée respectivement selon que l'on soit après la cassure ou avant la cassure.

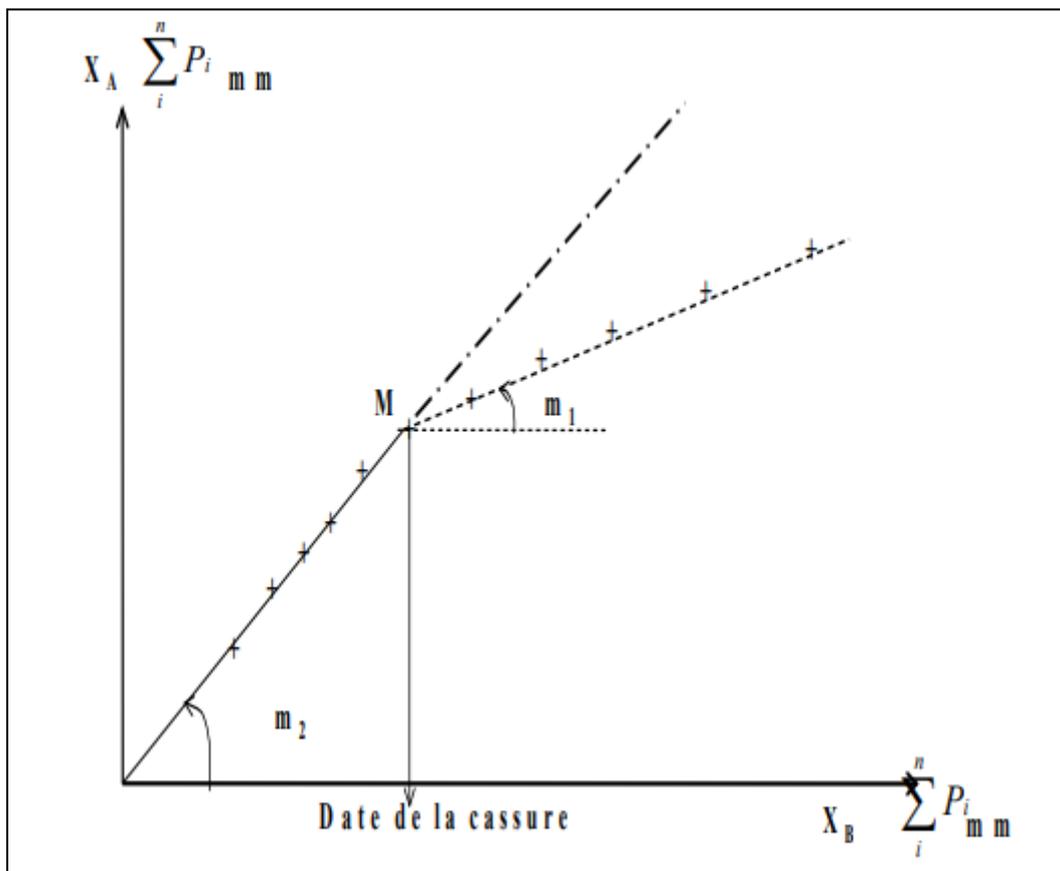


Figure 2.5 : Méthode des doubles masses

4.2.1.3 Méthode des résidus (Bois, 1972)

L'étude d'homogénéité des séries par la méthode du double cumul est généralement confrontée au manque de tests statistiques valables pour préciser la signification des cassures apparentes.

P. Bois a suggéré une méthode basée sur le cumul des résidus ϵ_i . Cette méthode permet de déterminer un seuil à partir duquel la cassure est significative.

Le problème revient à rechercher une courbe de contrôle telle que tout dépassement amène à repousser l'hypothèse d'homogénéité de la série avec un seuil de confiance Choisi. P Bois a montré que la courbe de contrôle est une ellipse d'équation:

$$Y_i = \pm t_{\alpha/2} \cdot \sigma_y \cdot \sqrt{1-r^2} \cdot \sqrt{\frac{i(n-i)}{n-1}} \quad 2.44$$

Où $\sigma_y \cdot \sqrt{1-r^2} \cdot \sqrt{\frac{i(n-i)}{n-1}}$ est la variance des résidus. Ces résidus étant déterminés par la relation suivante :

$$\varepsilon_i = y_i - \bar{y} - r \cdot \frac{\sigma_y}{\sigma_x} (x_i - \bar{x}) \quad 2.45$$

ε_i est donc la différence entre la valeur observée et la valeur donnée par la régression linéaire de y en x.

$t_{\alpha/2}$: Variable de Student ;

σ_y : L'écart type de y ;

σ_x : L'écart type de x ;

\bar{y} : Moyenne de y ;

\bar{x} : Moyenne de x ;

r_{xy} : Coefficient de corrélation entre les deux variables.

Cette méthode permet de dilater les différences, ce qui augmente considérablement la clarté du graphique surtout dans le cas où les écarts types sont faibles.

4.2.2 Tests analytiques

4.2.2.1 Test de la médiane

Soit un échantillon de n valeurs x_i (série chronologique par exemple) de médiane m. Chaque observation x_i est affectée d'un signe + si elle est supérieure à la médiane, d'un signe - dans le cas contraire. Tout groupe de valeurs « + » est une séquence positive (ou suite positive), et tout groupe de valeurs « - » est une séquence négative (ou suite négative).

Il s'agit de déterminer NS le nombre total de séquences positives ou négatives ainsi que TS, la taille de la plus longue de ces suites.

Ns suit approximativement une loi normale de moyenne $\frac{N+1}{2}$ et de variance $\frac{n-1}{4}$. Ts suit une loi binomiale.

Ceci a permis d'établir que pour un seuil de signification compris entre 91% et 95%, les conditions du test sont les suivantes :

$$\frac{1}{2} \left(N + 1 - u_{1-\frac{\alpha}{2}} \sqrt{N-1} \right) < Ns < \frac{1}{2} \left(N + 1 + u_{1-\frac{\alpha}{2}} \sqrt{N-1} \right) \quad 2.46$$

$$Ts < 3.3(\log_{10} N + 1) \quad 2.47$$

Si les conditions du test sont vérifiées, on conclut que la série à étudier est homogène au seuil de signification $1 - \alpha$.

4.2.2.2 Test de Wilcoxon ou Test des rangs

Soient 2 variables aléatoires Y et X, représentant respectivement 2 séries de taille N_1 et N_2 . Y étant la série à étudier et X étant la série de base avec $N_2 > N_1$. Si l'échantillon Y est issu de la même population que l'échantillon X, l'échantillon nouveau $Y \cup X$ est également issu de la même population. De ce fait, on classe les éléments de ce nouvel échantillon $Y \cup X$ par ordre croissant et on attribue à chacune des valeurs le rang qu'elle occupe dans cette nouvelle série. (Si une valeur se répète plusieurs fois, il faut lui associer le rang moyen qu'elle détermine). On calcule les quantités W_Y et W_X :

W_Y représente la somme des rangs de Y et c'est celle qui nous intéresse et elle est égale à :

$$W_Y = \sum_{i=1}^n rang_Y = 1 + 3 + 4 + \dots + 13 + 17 + \dots + n \quad 2.48$$

$$W_X = \sum_{i=1}^{n-1} rang_X = 2 + 5 + \dots + 12 + 14 + 15 + 16 + \dots + n - 1 \quad 2.49$$

L'hypothèse nulle est vérifiée si :

$$W_{min} < W_Y < W_{max}$$

Avec :

$$W_{min} = \frac{(N_1 + N_2 + 1)N_1 - 1}{2} - u_{1-\frac{\alpha}{2}} \sqrt{\frac{N_1 N_2 (N_1 + N_2 + 1)}{12}} \quad 2.50$$

$$W_{max} = (N_1 + N_2 + 1)N_1 - W_{min} \quad 2.51$$

$u_{1-\alpha/2}$ représente la valeur de la variable centrée réduite de Gauss correspondant à une probabilité de $1 - \alpha/2$.



Mise
en
application

Exercice 1

Soient 2 stations pluviométriques "Ain Bell" et "Ain Mahdi" situées à quelques kilomètres l'une de l'autre dans le bassin versant du Chot Melghir. Ces stations ayant fonctionné respectivement sur des périodes de 24 ans (N) et de 22 ans (K) comme le montre le tableau 2.3.

En supposant que la série pluviométrique des précipitations annuelles de la station de Ain Bell est la station de référence (X) et que l'erreur recherchée se trouve au niveau de la série pluviométrique de Ain Mahdi, série à étudier (Y), on demande de :

- 1- Vérifier l'homogénéité de la série de la station de référence (X) en appliquant le test de la médiane ;
- 2- Détecter l'erreur systématique de la station étudiée et faire la correction par la méthode des doubles masses s'il y a erreurs ;
- 3- Vérifier l'homogénéité de la série Y en appliquant le test de Wilcoxon ;
- 4- Donner la droite de régression de Y en X ;

Tableau 2.3 : Données pluviométriques

Année	Ain Bell (X)	Ain Mehdi (Y)	Année	Ain Bell (X)	Ain Mehdi (Y)
1980	348,1	213,7	1992	140	70,5
1981	295	229	1993	120	81,5
1982	208,1	106,3	1994	386,8	111,2
1983	50,7	25,1	1995	390,9	266,3
1984	225,7	102	1996	119,4	78,1
1985	350,4	161,9	1997	312,3	190
1986	186,6	141,4	1998	69,3	106,3
1987	221,9	86	1999	175,4	145,8
1988	216,8	174,4	2000	95,3	68,9
1989	446,3	208,1	2001	148,9	91,9
1990	213,7	195,6	2002	131,8	lac
1991	351,8	186,7	2003	310,5	lac

Solution

1. Test de la médiane : série de référence ou de base

Le test de la médiane (test de Mood) étant réalisé, l'homogénéité étant vérifiée, ce test permettra de voir si la série à étudier est homogène ou pas, c'est-à-dire si elle appartient à la même population que la série de référence.

La médiane m déterminée sur la série Y classée par ordre croissant est : **m= 215.25 mm.**

L'application du test nécessite la vérification des conditions Ns et Ts (Tableau 5.1).

Tableau 2.4 : Test de la médiane (série de référence)

N°	Année	Px (mm)		N°	Année	Px (mm)	
1	1980	348,1	+	13	1992	140	-
2	1981	295	+	14	1993	120	-
3	1982	208,1	-	15	1994	386,8	+
4	1983	50,7	-	16	1995	390,9	+
5	1984	225,7	+	17	1996	119,4	-
6	1985	350,4	+	T_s 18	1997	312,3	+
7	1986	186,6	+	19	1998	69,3	-
8	1987	221,9	+	20	1999	175,4	+
9	1988	216,8	+	21	2000	95,3	-
10	1989	446,3	+	22	2001	148,9	-
11	1990	213,7	-	23	2002	131,8	-
12	1991	351,8	+	24	2003	310,5	+

Pour : $N=24$ et $u_{1-\alpha/2} = 1.96$ (variable de Gauss, lu sur la table de Gauss pour un seuil de signification $1-\alpha = 95\%$).

$$\frac{1}{2} \left(N + 1 - u_{1-\frac{\alpha}{2}} \sqrt{N-1} \right) < N_s < \frac{1}{2} \left(N + 1 + u_{1-\frac{\alpha}{2}} \sqrt{N-1} \right)$$

$N_s = 13$ (nombre total de séries de + ou de - dans la série initiale)

$$T_s < 3.3(\log_{10} N + 1) \rightarrow T_s > 4.61$$

$T_s=6$ (taille de la plus grande série de + ou de - au-dessus de la médiane dans la série initiale).

On déduit que :

$$\rightarrow 7.80 < N_s < 34.39$$

Pour T_s : On a : $6 < 4.61$

} La série de référence est homogène

L'homogénéité de la série de référence étant vérifiée, cette série servira de base pour la détection d'erreurs systématiques dans la série à étudier.

2. Test des doubles cumuls

La station de référence est celle de Ain Bell (X). Elle a été vérifiée par le test de la médiane. La station à étudier est celle de Ain Mahdi (Y) qu'il convient de vérifier, de corriger en cas d'erreurs et d'étendre pour son utilisation future. La méthode des doubles masses est considérée.

Le tableau 2.5 représente les valeurs initiales et cumulées des précipitations annuelles aux 2 stations pluviométriques. La méthode de la double masse appliquée aux cumuls annuels des 2

stations a permis de confirmer l'hétérogénéité de la série des pluies annuelles de la station Y comme le montre la figure 2.6.

Au vue de la figure 5.2, la station Y présente une hétérogénéité qu'il convient de corriger. Le changement ou la cassure de la pente correspond à l'année 1994. A partir de cette année, les 7 autres années qui suivent sont erronées et doivent être rectifiées.

Tableau 2.5 : test des doubles cumuls

Année	Px(mm)	Py(mm)	Cumul (X)	Cumul(Y)
1980	348,1	213,7	348,1	213,7
1981	295	229	643,1	442,7
1982	208,1	106,3	851,2	549
1983	50,7	25,1	901,9	574,1
1984	225,7	102	1127,6	676,1
1985	350,4	161,9	1478	838
1986	186,6	141,4	1664,6	979,4
1987	221,9	86	1886,5	1065,4
1988	216,8	174,4	2103,3	1239,8
1989	446,3	208,1	2549,6	1447,9
1990	213,7	195,6	2763,3	1643,5
1991	351,8	186,7	3115,1	1830,2
1992	140	70,5	3255,1	1900,7
1993	120	81,5	3375,1	1982,2
1994	386,8	111,2	3761,9	2093,4
1995	390,9	266,3	4152,8	2359,7
1996	119,4	78,1	4272,2	2437,8
1997	312,3	190	4584,5	2627,8
1998	69,3	106,3	4653,8	2734,1
1999	175,4	145,8	4829,2	2879,9
2000	95,3	68,9	4924,5	2948,8
2001	148,9	91,9	5073,4	3040,7

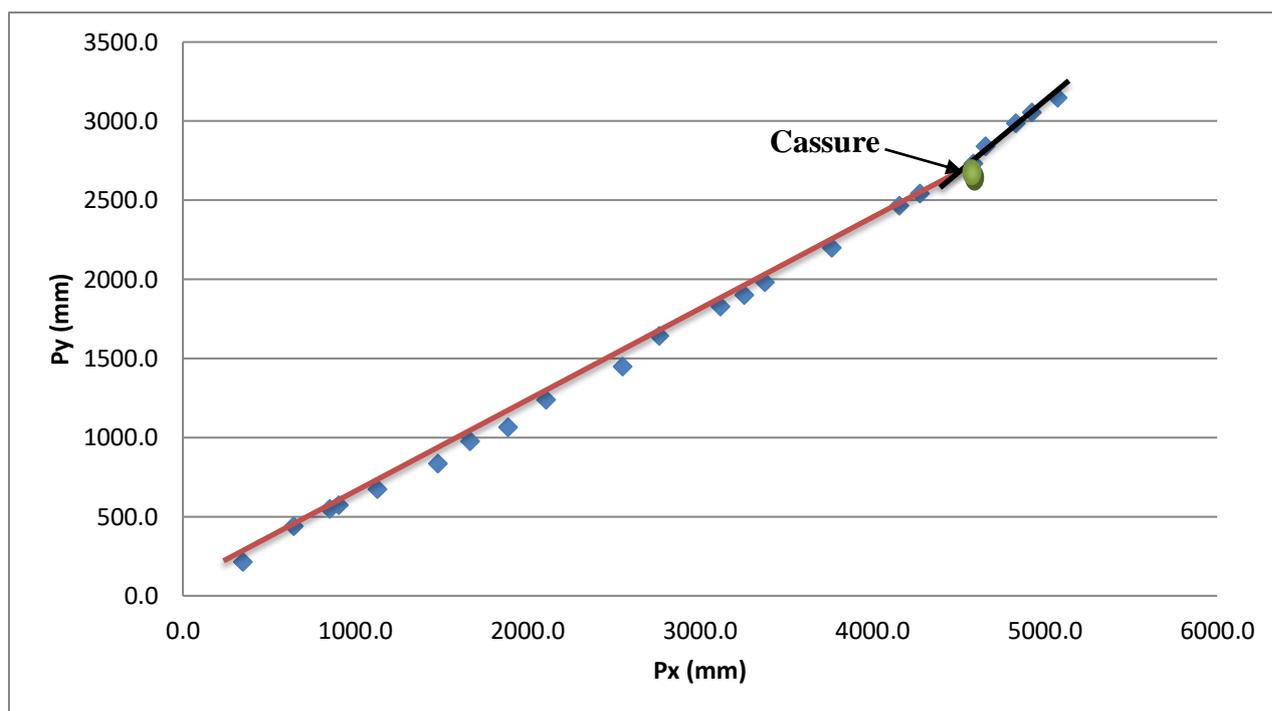


Figure 2.6 : test des doubles cumuls

Les pentes m_1 et m_2 correspondant respectivement aux 1er et 2ème segments de droite sont calculées : $m_1 = 1.61$; $m_2 = 1.28$.

On corrige les 5 dernières années de la station Y par un coefficient multiplicatif (rapport $m_1/m_2 = 1.36$).

Tableau 2.5 : Valeurs annuelles initiales et corrigées (Station Y)

Année	Pan (Y initiale) mm	Pan (Y corrigée) mm
1995	502,8	305,9
1996	111,6	171,1
1997	282,4	234,7
1998	153,4	110,9
1999	239,7	148,0
2000	502,8	305,9
2001	111,6	171,1

3. Test de Wilcoxon ou Test des rangs

Avant de faire l'extension, il convient de tester si la série corrigée appartient à la même population que la série de référence. Le test de Wilcoxon est le plus puissant des tests non paramétriques.

Tableau 2.6 : Test de Wilcoxon

Rang	X U Y	Rang Y	Rang X	Somme rang Y	Somme rang X
1	25,1	1		1	
2	50,7		2		2
3	68,9	3		4	
4	69,3		4		6
5	70,5	5		9	
6	78,1	6		15	
7	81,5	7		22	
8	86	8		30	
9	91,9	9		39	
10	95,3		10		16
11	102	11		50	
12	106,3	12		62	
13	106,3	13		75	
14	111,2	14		89	
15	119,4		15		31
16	120		16		16
17	131,8		17		17
18	140		18		18
19	141,4	19		108	
20	145,8	20		128	
21	148,9		21		39
22	161,9	22		150	
23	174,4	23		173	
24	175,4		24		63
25	186,6		25		25
26	186,7	26		199	
27	190	27		226	
28	195,6	28		254	
29	208,1	29		283	
30	208,1		30		55
31	213,7		31		31
32	213,7	32		315	
33	216,8		33		64
34	221,9		34		34
35	225,7		35		99
36	229	36		351	
37	266,3	37		388	
38	295		38		137
39	310,5		39		176
40	312,3		40		216
41	348,1		41		257
42	350,4		42		299
43	351,8		43		342
44	386,8		44		386
45	390,9		45		431
46	446,3		46		477

NB : les cellules en jaune représentent les valeurs de la série à vérifier (X).

$$W_Y = \sum_{i=1}^n rang_Y = 388$$

$$W_X = \sum_{i=1}^{n-1} rang_X = 477$$

$$W_{min} = \frac{(N_1 + N_2 + 1)N_1 - 1}{2} - u_{1-\frac{\alpha}{2}} \sqrt{\frac{N_1 N_2 (N_1 + N_2 + 1)}{12}} = 172.3$$

$$W_{max} = (N_1 + N_2 + 1)N_1 - W_{min} = 955.7$$

La condition du test vérifiée $172.3 < W_Y < 955.7$. Les 2 séries appartiennent à la même population et elles sont donc homogènes.

4. Droite de régression

Après correction des valeurs erronées de la station Y, la relation est plus nette. La droite de régression régnant la relation Y en X est

$$Y = 0.426X + 39.9$$

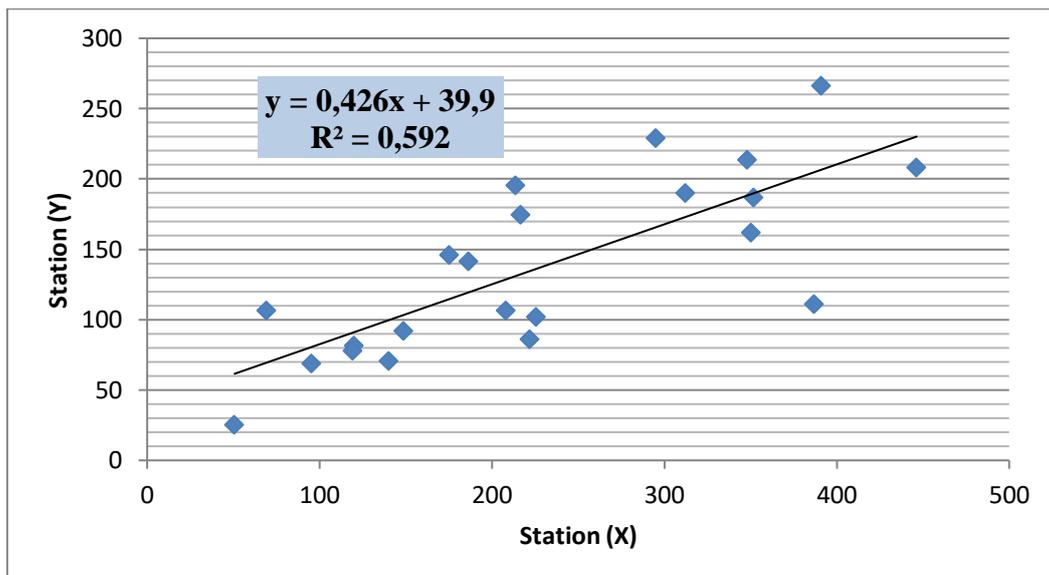


Figure 2.7 : La droite de régression

Exercice 2

L'objectif de l'exercice est d'initier les étudiants au logiciel Khronostat.

Afin d'étudier la composante de la tendance centrale, 05 stations pluviométriques ont été utilisées pour une superficie de 3677.5 km².

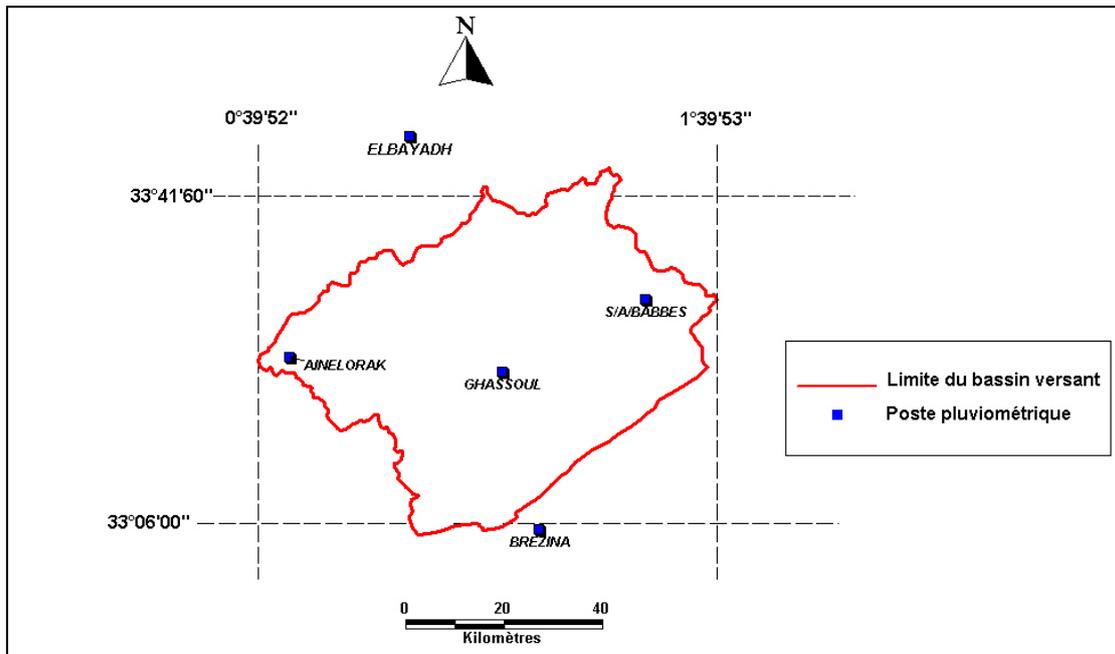


Figure 2.8: Répartition des postes pluviométriques

1. Appliquer la méthode de la moyenne mobile.
2. Etudier le caractère aléatoire des séries chronologiques.
3. Appliquer les tests de détection de rupture.

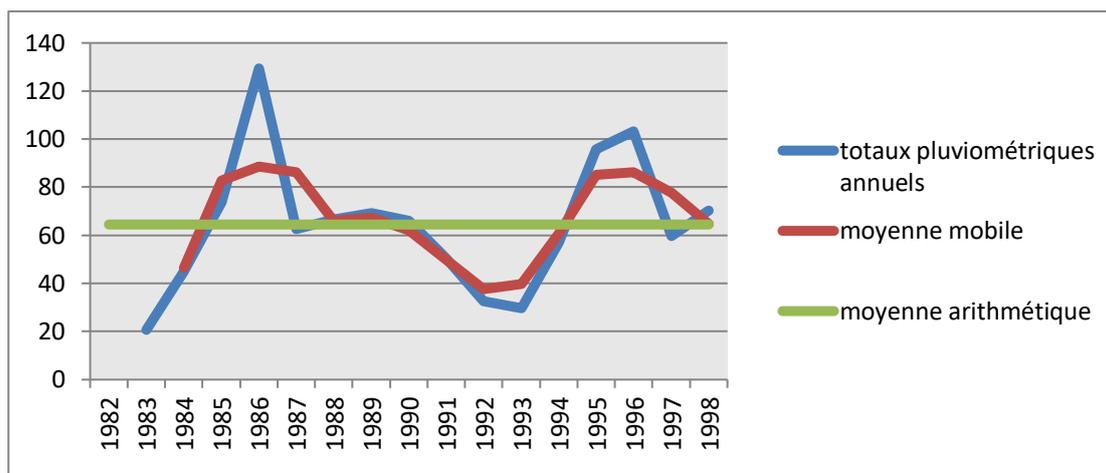
Solution

1. Méthode de la moyenne mobile

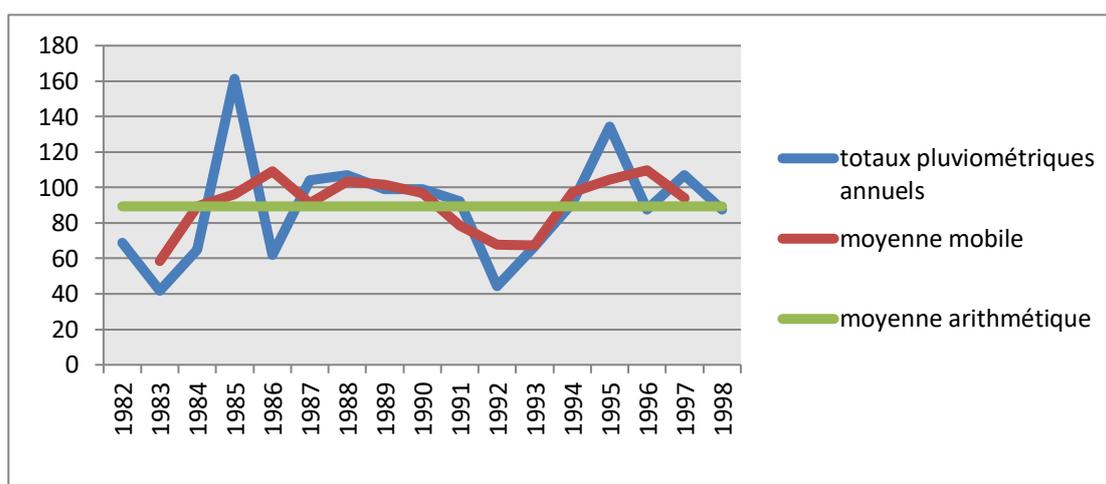
Chaque observation est remplacée par une moyenne arithmétique calculée sur la valeur de cette observation et les valeurs voisines qui l'encadrent ; dans le cadre de notre étude on a utilisé trois valeurs : ainsi x_i est remplacé dans la série par :

$$\frac{x_{i-1} + x_i + x_{i+1}}{3} \dots\dots\dots \text{(Grisolet H ; 1972)}$$

Cette méthode réduit l'influence des variations accidentelles, élimine l'effet des fluctuations de très courtes périodes et par conséquent les résultats sont plus commodes à interpréter. On a opté pour cette méthode en raison de sa simplicité.



**Figure 2.9 : Variation annuelle des précipitations moyennes annuelles (1982/1998)
-Station Ain Lorak-**



**Figure 2.10: Variation annuelle des précipitations moyennes annuelles (1982/1998)
-Station S/A/Babbes-**

A l'Ouest et à l'Est du bassin versant, l'étude de variabilité interannuelle sur la période 1982-1998 pour les stations d'Ain Lorak et S/B/ Babbes (la figure 2.9 et la figure 2.10), montre que le plus grand nombre d'années déficitaires s'observe durant le début de la décennie 90. Ces résultats montrent également la persistance des totaux déficitaires qui se sont prolongés sur plusieurs années successives. Entre cette grande période de sécheresse, la pluviométrie a été généralement normale ou excédentaire, on a même enregistré 6 années pluvieuses consécutives à la fin des années 90 au niveau de la station d'Ain Lorak. L'année la plus déficitaire, pendant la période d'étude est observée en 1993 au niveau des deux stations, durant cette année, le total pluviométrique annuel n'a atteint que 29,7 mm à la station d'Ain Lorak.

Dans le Centre et à l'extrême sud au niveau de les deux stations Ghassoul et Brezina (La Figure 2.11 et la Figure 2.12) les variations ne sont pas semblables aux deux stations précédentes. La première différence concerne la première tendance à la hausse qui a débuté la fin des années 80 et s'est prolongée jusqu'au début des années 90. À partir de cette date jusqu'au milieu des années 90, on observe une tendance à la baisse.

2. la vérification du caractère aléatoire

Nous avons utilisé le test de corrélation sur le rang et l'autocorrélogramme pour vérifier le caractère aléatoire des séries chronologiques. Le tableau 2.7 montre que ces séries sont des composantes aléatoires, non systématiques, irrégulières, c'est-à-dire dues au hasard. Cette constatation est appliquée ainsi aux séries chronologiques saisonnières et mensuelles.

La méthode de l'auto- corrélogramme affirme ces constatations. Voir la figure 2.11.

Tableau 2.7 : Les résultats de l'application du test de corrélation sur le rang

station	Valeur de la variable de calcul (U)	Seuil de confiance de 90%	Seuil de confiance de 95%	Seuil de confiance de 99%
Ain Lorak	0.6303	Acceptée	Acceptée	Acceptée
Ghassoul	1.4005	Acceptée	Acceptée	Acceptée
S/A/Babbes	0.4943	Acceptée	Acceptée	Acceptée
Brezina	-0.1801	Acceptée	Acceptée	Acceptée
El Bayadh	-0.1648	Acceptée	Acceptée	Acceptée

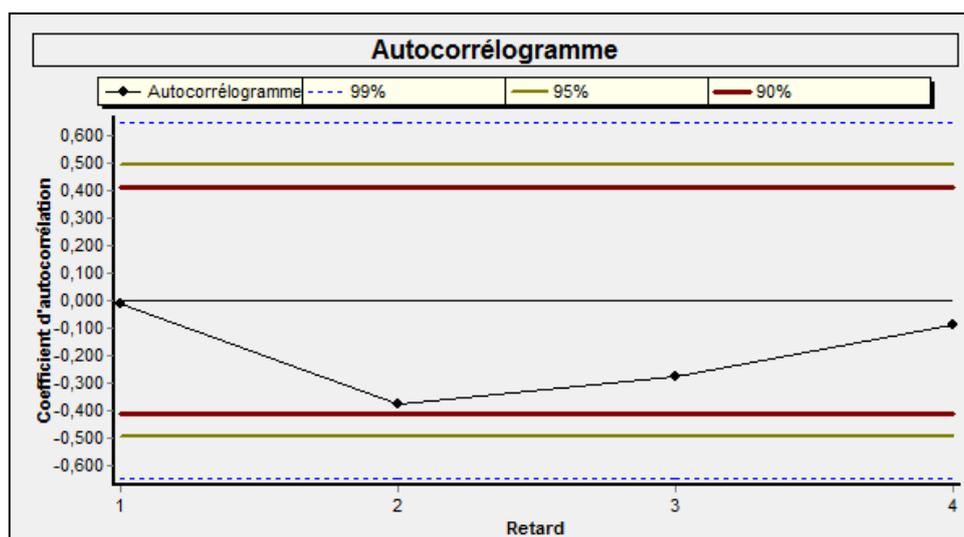


Figure 2.11: Autocorrélogramme -Station Brezina-

3. Détection des ruptures dans les totaux pluviométriques annuels :

L'application des tests de détection de rupture (Test de Pettitt et la segmentation d'Hubert) aux séries pluviométriques annuelles est représentée dans le tableau 2.8 et le tableau 2.9. Nous avons constaté qu'au niveau de la zone d'étude, aucune rupture significative n'a pu être détectée et ceux malgré la diminution des totaux annuels lors des deux dernières décennies.

Tableau 2.8 : Les résultats du test de Pettitt.

Station	Intervalle de confiance à 90%	Intervalle de confiance à 95%	Intervalle de confiance à 99%
Ain Lorak	Acceptée	Acceptée	Acceptée
Ghassoul	Acceptée	Acceptée	Acceptée
S/A/Babbes	Acceptée	Acceptée	Acceptée
Brezina	Acceptée	Acceptée	Acceptée
El Bayadh	Acceptée	Acceptée	Acceptée

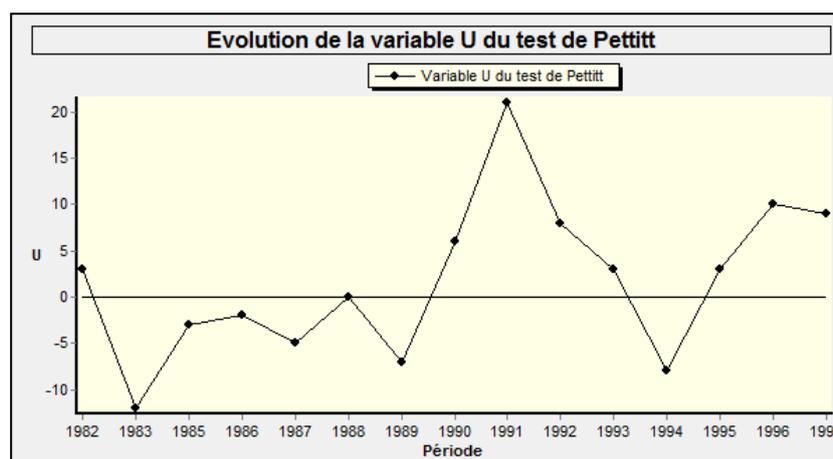


Figure 2.13: Évolution de la variable U du test de Pettitt pour -Station Brezina-

Tableau 2. : Les résultats de la Méthode de Segmentation d'Hubert.

Station	Début	Fin	Moyenne	Ecart type
Ain Lorak	1982	1997	64.500	27.859
Ghassoul	1982	1993	59.558	18.392
	1994	1998	90.720	22.097
S/A/Babbes	1982	1998	89.353	30.445
Brezina	1982	1998	66.706	42.647
El Bayadh	1982	1998	115.018	28.708

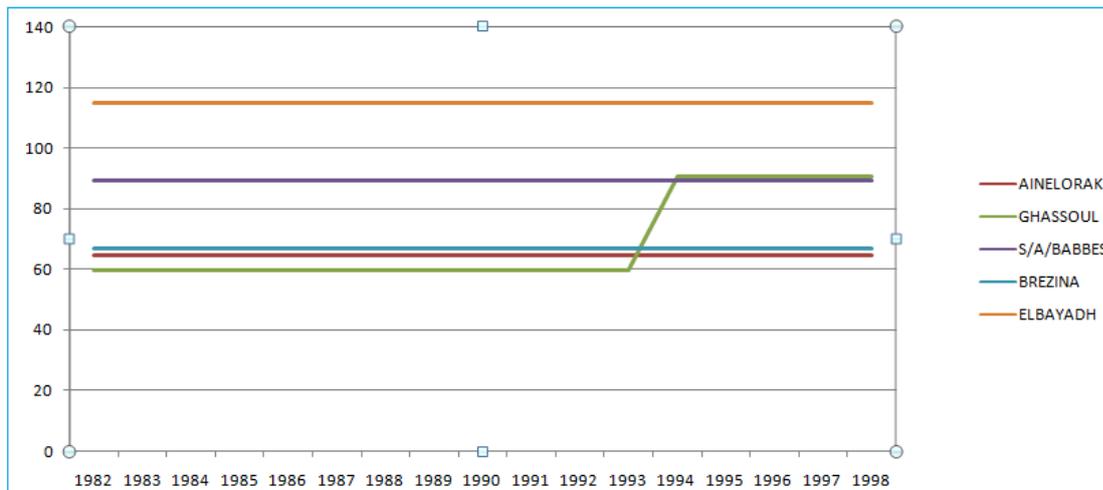


Figure 2.14 : Les résultats de la Méthode de Segmentation d'Hubert (1982 à 1998).

Chapitre 3 : Modélisation hydrologique

Objectif du chapitre

En hydrologie, une représentation mathématique simplifiée de tout ou une partie des processus du cycle hydrologique est indispensable. Donc les concepts hydrologiques sont exprimés en langage mathématique pour représenter le comportement correspondant observé dans la nature. L'intérêt d'un modèle réside dans sa capacité à apporter une réponse "satisfaisante" aux questions qu'il se pose à propos de l'objet modélisé.

1. La modélisation hydrologique

1.1 Définition du modèle hydrologique

Un modèle hydrologique est donc, un ensemble d'équations mathématiques qui représentent l'état d'un bassin versant et l'évolution du cycle de l'eau par rapport à des conditions météorologiques spécifiées. Les éléments du cycle de l'eau modélisés sont les processus hydrologiques (infiltration, ruissellement, évaporation, évapotranspiration, fonte de la neige, ...). Il s'agit donc d'une représentation plus ou moins simplifiée du bassin versant et des diverses composantes du cycle hydrologique (partiel ou complet).

1.2. Quelques éléments de vocabulaire

1.2.1. Les variables d'un modèle hydrologique

Un modèle hydrologique est généralement défini par (Figure 3.1):

- Variable d'entrées : le modèle fait appel à ces variables qui dépendent du temps et/ou de l'espace (pluie, ETP, caractéristiques physiques et hydrodynamique du milieu).
- Variable de sortie : le modèle répond par un ensemble de variables (débits, flux ou concentration en polluant,)
- Variable d'état : il s'agit des variables internes au système, qui évolue en fonction du temps et rendent compte de l'état du système à un moment donné. Typiquement, ces variables sont les niveaux de remplissage des différents réservoirs (neige / production / routage).

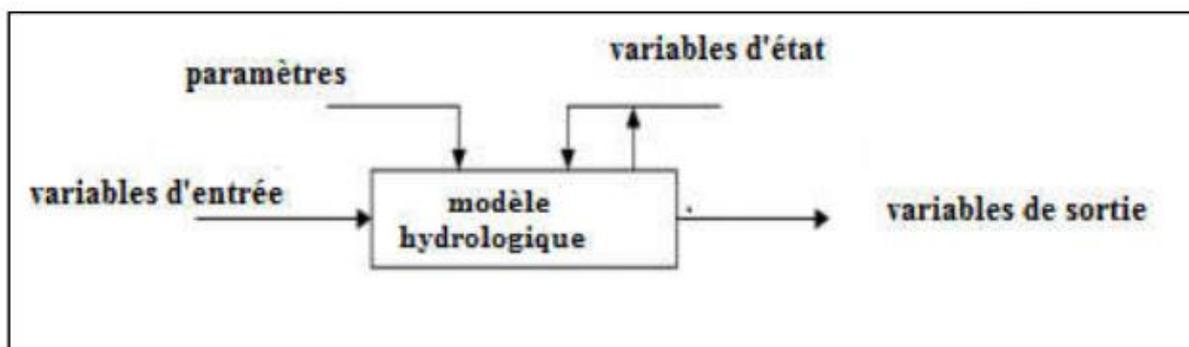


Figure 3.1 : Les variables d'un modèle hydrologique (Gaume, 2002)

Les variables citées ci-dessus interviennent dans la modélisation hydrologique par l'intermédiaire de deux fonctions : une fonction de production et une fonction de transfert:

La fonction de production : c'est une représentation simple mais réaliste, des différentes voies que suivra l'eau de pluie, entre le moment où elle rejoint le cours d'eau (Morin, 1991 ; Boudahraa, 2007), elle exprime la transformation de la pluie brute en pluie nette, définit la fraction de la pluie brute qui contribue effectivement au ruissellement. En d'autres termes, elle permet de calculer la quantité d'eau qui va s'écouler à l'exutoire d'un bassin ou sous bassin versant.

La fonction de transfert : c'est la fonction qui permet de transférer, comme son nom l'indique, la quantité d'eau déterminée par la fonction de production, vers le cours d'eau. Elle permet la transformation de la pluie nette en un hydrogramme à l'exutoire du bassin versant, donc, de donner une forme à la crue dont le volume a été déterminé par la fonction de production, en simulant l'hydrogramme de crue à l'exutoire.

2. Pourquoi des modèles hydrologiques ?

La modélisation hydrologique définit les relations quantitatives entre les caractéristiques de l'écoulement (sorties) et les facteurs influant sur ces valeurs (entrées). C'est une définition très générale qui englobe un large éventail de méthodes.

Dans ce chapitre, on se limite à l'étude des modèles pluie-débit, qui représentent la transformation de la pluie en écoulement. Les applications de ces modèles sont multiples, et permettent la :

- Reconstitution des séries de débit par le comblement de lacunes dans les séries historiques afin de permettre les traitements statistiques utilisés notamment dans les études de faisabilité des ouvrages et de détermination des valeurs de projet;
- Prévision des fréquences avec lesquelles les débits extrêmes (crue et étiage) peuvent se produire ;
- Prédétermination de la réponse hydrologique d'un bassin versant avec un pas de temps, afin de permettre l'élaboration d'un système de prévention et d'alarme ;
- Etude de l'impacte des différents aménagements sur la réponse hydrologique.

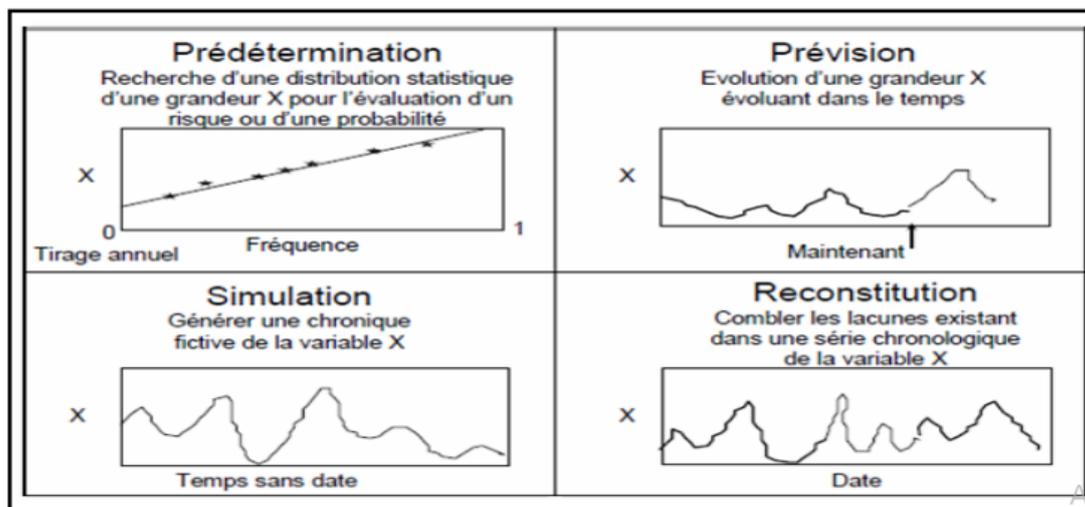


Figure 3.2 : Les opérations de la modélisation

3. Différentes approches de modélisation

Les modèles hydrologiques sont des représentations simplifiées du cycle de l'eau. Un large éventail de types de modèles hydrologiques est disponible pour l'utilisation par les scientifiques et les ingénieurs des processus hydrologiques sont des modèles à l'échelle et des modèles mathématiques, un modèle à l'échelle est une réplique miniaturisée d'un système physique réel et est composé de matériaux similaires à ceux que l'on trouve dans le système réel étudié (Thompson, 2017). Ambroise (1998) présente une classification des modèles hydrologiques basée sur les modèles conceptuels et les modèles « physico-conceptuels », cette classification permet de pallier l'ambiguïté entre les classes de modèles conceptuels et à base physique. En effet, le terme de modèle conceptuel reprend des modèles très éloignés dans leur fonctionnement : celui-ci peut être physiquement basé ou non, le terme conceptuel ne faisant référence alors qu'à une représentation schématique du bassin versant en réservoirs (Gineste, 1997). Les modèles conceptuels à base physique ou capacitifs. Beckers and Degré (2011) donne une description des différents modèles hydrologiques selon la classification suivante : les modèles empiriques ou « boîtes noires » ; les modèles conceptuels empiriques ou « boîtes grises » ; les modèles conceptuels à base physique ou capacitifs et les modèles physiques.

Depuis la révolution informatique, la modélisation hydrologique a fait un énorme bond en avant, ce qui a donné naissance à une nouvelle branche de l'hydrologie, appelée hydrologie numérique ou digitale (Singh, 2018). Actuellement, les modèles hydrologiques sont des outils indispensables et largement utilisés pour la prévision des variables hydrologiques, gestion de réservoirs, prise de décision ou encore pour améliorer la compréhension des processus régissant le cycle hydrologique d'un bassin versant. Ces outils sont des ensembles d'équations mathématiques qui décrivent et représentent de manière relativement simple les processus hydrologiques occurrence dans un bassin versant (Larabi, 2017). Un modèle mathématique peut être défini comme un ensemble d'équations mathématiques combinées avec des énoncés logiques qui simulent le fonctionnement d'un système réel (Refsgaard, 1997). Étant donné que la discussion de tous les types de modèles hydrologiques, ou même des deux, dépasse le cadre de ce chapitre, la présente analyse documentaire se concentre exclusivement sur le groupe mathématique des modèles hydrologiques. Les modèles hydrologiques visent la simplicité en sélectionnant les aspects fondamentaux du système cible au détriment des détails accessoires (Allen et Freeze, 1987). Un certain nombre de techniques alternatives et d'approches de modélisation ont été développées. Le premier modèle hydrologique intégré, appelé modèle de captage de Stanford, a été publié dans la littérature en 1966 par Crawford et Linsley (Singh and Frevert, 2002). Au cours des décennies suivantes, la modélisation hydrologique s'est considérablement améliorée grâce aux progrès de la technologie et du l'outil informatique. De meilleurs modèles hydrologiques deviennent disponibles grâce à ces progrès technologiques et à l'amélioration continue des techniques de modélisation, comme les SIG, la télédétection ou les automates cellulaires.

Selon Singh, 1995 (figure 3.3), un modèle est caractérisé par 5 éléments constitutifs : la géométrie du système (bassin versant), les entrées dans le système, les lois de formalisations des processus, l'état initial et les conditions aux limites du système et enfin les sorties. Singh (1995) a proposé trois critères pour classer les modèles hydrologiques qui sont : la description des processus, l'échelle spatio-temporelle et la méthode utilisée pour résoudre les équations.

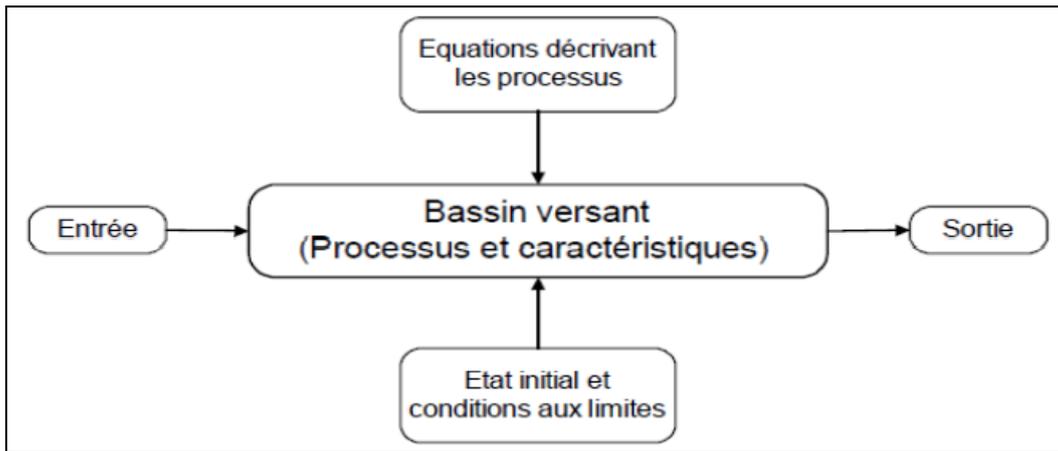


Figure 3.3 : Composants d'un modèle (Singh et al, 1995)

Les approches habituellement utilisées dans la modélisation pluie-débit apparaissent dans la Figure 3.4.

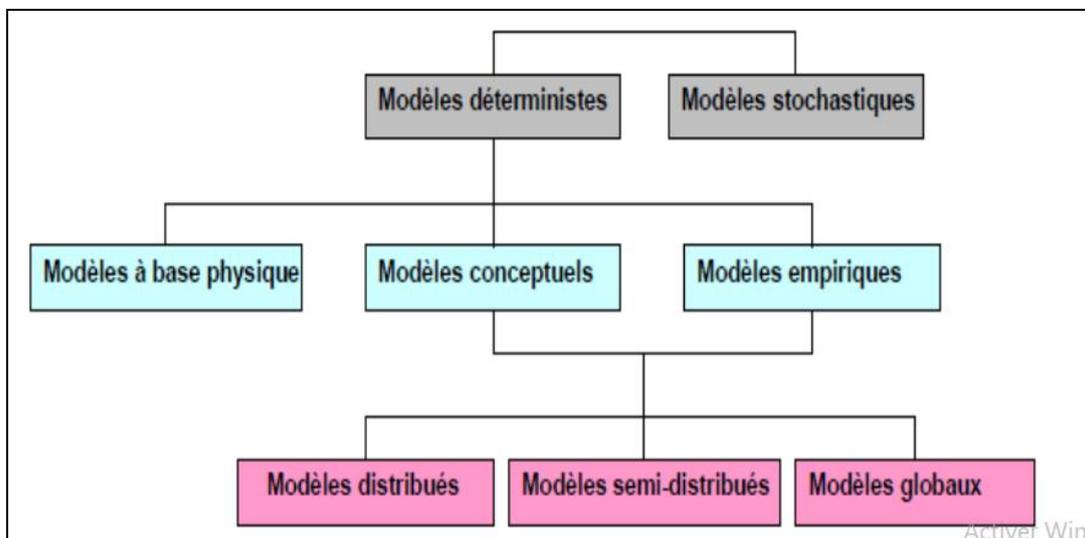


Figure 3.4: Relation entre les modèles pluie-débit (Berthier, 2005)

3.1. Selon le degré d'abstraction Modèles physiques

Dont les équations ont été déduites à partir des principes de base de la physique (conservation de masse, quantité de mouvement, quantité d'énergie, ...), et dont la forme finale simplifiée contient des paramètres qui ont un sens physique. Ils représentent le système à une échelle réduite. Il existe plusieurs modèles à bases physiques comme le modèle KINEROS, LISEM, ABC,...

Modèles mathématiques : Ils décrivent le processus hydrologique à l'aide des équations mathématiques qui relient les variables d'entrée et de sortie.

3.2. Selon la nature des variables

- **Modèles déterministes** : Dans lesquels la relation entre variables d'entrée et de sortie est phénoménologique et ne considère pas le caractère aléatoire de variable, la même entrée produit toujours la même sortie.

- **Modèles stochastiques (probabilistes)** : Le terme stochastique est lié à la statistique et traduit le caractère aléatoire du modèle. Il permet de générer, aléatoirement, des données par des lois de distributions particulières.

3.3. Selon la discrétisation spatiale

- **Modèles globaux** : D'après Maison (2000), le modèle global s'applique globalement sur l'ensemble d'un bassin versant (approche par bilan) qui est considéré comme entité homogène. On utilise dans ce type de modèles des valeurs représentatives moyennes. Un exemple de ce type de modèle est le modèle GR (génie rural) (CEMAGRAF) que nous allons utiliser dans notre étude. C'est un modèle à réservoir, il fonctionne pour différents pas de temps annuel GR1A, mensuel GR2M et journalier GR4J.

- **Modèles spatialisés (distribués)** : Ils tiennent en compte la variabilité spatiale de processus et de variables d'entrée, ils sont plus avantageux que les modèles globaux. Dans le cas où les composants du système ne sont pas complètement spatialisés, on parle de modèles semi spatialisés.

3.4. Selon la description des sous processus

Le processus hydrologique de transformation pluie-débit est composé d'autres sousprocessus : précipitations, interception, infiltration, ... etc. Les modèles hydrologiques peuvent se classer selon la manière de décrire ces processus :

- **Modèles conceptuels** : D'après Ambroise (1991), le modèle conceptuel considère le bassin versant comme un assemblage de réservoirs d'humidité, interconnectés et qui sont censés de représenter plusieurs niveaux de stockages, suivant une dimension verticale. Ce modèle est basé sur la connaissance des phénomènes physiques qui agissent sur les entrées pour obtenir les sorties.

- **Modèles empiriques** : Ils sont utilisés pour reproduire le comportement global du système sans décrire les processus élémentaires. Ils sont généralement de type boîte noire sous la forme débit = f (pluie) où f est déterminée par une régression effectuée sur les chroniques pluies-débits (Ambroise, 1991).

3.5. Choix d'un modèle

Le choix d'un modèle hydrologique se fait en fonction des éléments suivants :

- objectif de l'étude hydrologique : Les objectifs de l'étude hydrologique sont multiples selon lesquels on peut négliger ou simplifier certains paramètres dans la modélisation ;

- disponibilité de données : En dépit des incertitudes que présentent des modèles par rapport à d'autres, le manque de données nous contraint à les utiliser ;

- nature de modèle, la robustesse et la simplicité de modèle influencent le choix en particulier dans le cas où le temps ou le coût présentent une contrainte.

3.6. Phase de calage et de validation d'un modèle

3.6.1. Calage

Le calage d'un modèle consiste à déterminer le jeu de paramètres du modèle permettant de reproduire le plus fidèlement possible les débits observés à l'exutoire. Lors du calage, le problème qui se pose est le nombre de données à caler, à ce propos il n'existe pas une règle commune entre tous les bassins, en générale, la calibration se fait sur les deux tiers de la série de base. (Chiheb, 2011) Quelque soit l'approche utilisée, le calage du modèle exige (Maklhouf, 1994) :

- Une série de données "représentative" des différentes entrées dans le système,
- Une estimation initiale des paramètres et des variables d'état du système,
- Une méthode d'ajustement des paramètres,
- Une méthode ou un critère permettant d'évaluer la qualité de la simulation des débits par le modèle.

3.6.2. Méthode de calage des modèles pluie-débit

On peut procéder au calage de modèle par trois méthodes :

- Calage manuel Cette méthode consiste à donner des valeurs aux paramètres du modèle et évaluer manuellement l'erreur entre les valeurs de sortie et les valeurs observées et l'échantillon de référence pour faciliter cette opération on à une représentation graphique.
- Calage automatique Ce calage optimise automatiquement les paramètres de modèle par l'utilisation d'un algorithme numérique pour trouver un extremum d'un critère numérique donné afin de déterminer l'ensemble des paramètres qui satisferont un critère donné de précision à travers différentes combinaisons possibles (Kingumbi, 2006).
- Calage mixte Le calage mixte consiste à combiner les deux méthodes précédentes. On détermine manuellement l'intervalle de variation des paramètres, puis on utilise la méthode automatique pour trouver les valeurs de paramètres optimales.

3.6.3. Validation

Cette période permet de vérifier la robustesse du calage en terme d'indépendance du jeu de paramètres vis à vis des données d'entrées et donc l'aptitude du modèle à fournir des simulations (ou des prévisions) dans diverses situations hydrologiques (Quesney, 1999). Cependant, il est important de distinguer le calage de la validation, la première opération assurant l'adéquation des paramètres alors que la seconde évalue l'adéquation du modèle tout entier.

3.6.4. Critères d'appréciation des modèles

La calibration d'un modèle hydrologique nécessite de définir un ou plusieurs critères de performance. Les critères possibles sont multiples. Peuvent être aussi utilisés pour l'étape d'évaluation du modèle. La performance d'un modèle peut être évaluée qualitativement en comparant visuellement, sur la base d'une même représentation graphique, les valeurs observées et celles simulées d'une variable donnée. Plusieurs types de représentations graphiques peuvent être considérés. Il peut s'agir, par exemple de :

- L'évolution temporelle, à différents pas de temps et sur une fenêtre temporelle donnée, de la variable d'intérêt.
- La structure spatiale, à un instant donné, de la variable d'intérêt si le modèle est spatialisé, semi-spatialisé et si les observations correspondantes sont disponibles.
- La distribution statistique de la variable d'intérêt (distribution statistique des débits journaliers, des débits maximum de crue au pas de temps horaire).

Pour tester la robustesse du modèle nous devons calculer des critères qui estiment la différence entre les données observées et les données simulées. On distingue deux types de critères (Benkaci, 2006):

- **Les critères graphiques** : correspondent à un ajustement graphique des données simulées sur les données observées. Ces critères ne sont pas sans importance, du fait qu'ils permettent un jugement visuel, et donc d'apprécier la qualité du modèle à reproduire les différentes tendances climatiques, tels que la succession des cycles de dessèchement et de rehumectation par le modèle.

- **Les critères numériques** : consistent à déterminer une fonction qui va minimiser l'erreur entre les données simulées par le modèle, et celles observées. On peut citer :

a. L'erreur quadratique (E)

C'est le critère le plus utilisé, il consiste à minimiser la somme des carrés des erreurs entre les débits observés et les débits calculés, il permet d'évaluer l'erreur entre la variance et le biais (Benkaci, 2006) :

$$E = \sum_{i=1}^n (Q_{i\text{ obs}} - Q_{i\text{ sim}})^2 \quad (3.1)$$

Où :

$Q_{i\text{ obs}}$ = Débit observé au pas de temps i ;

$Q_{i\text{ sim}}$ = Débit simulé au pas de temps i ;

n = Taille de l'échantillon.

b. Le critère RMSE (Root Mean Squared Error)

Il est défini comme étant la racine carrée de l'erreur quadratique, l'introduction de la racine carrée a pour but d'homogénéiser les écarts entre les débits simulés et observés en particulier dans les cas extrêmes (débits faibles ou élevés), du fait que cette erreur est élevée au carré dans le cas du critère quadratique. Le critère RMSE exprimé en m³/s est défini comme suit :

$$RMSE = \sqrt{\sum_{i=1}^n (Q_{i\text{ obs}} - Q_{i\text{ sim}})^2} \quad (3.2)$$

c. RMSE-observations standard déviation ratio (RSR)

Le RSR est l'indice d'erreur le plus utilisé (Chu and Shirmohammadi, 2004; Singh et al, 2004; Vasquez-Amabile and Engel, 2005). Le modèle est plus performant quand la valeur du RMSE est petite.

$$RSR = \frac{RMSR}{STDV_{obs}} = \frac{\sqrt{\sum_{i=1}^n (Q_{i\text{ obs}} - Q_{i\text{ sim}})^2}}{\sqrt{\sum_{i=1}^n (Q_{i\text{ obs}} - Q_{moy})^2}} \quad (3.3)$$

Le RSR incorpore les avantages d'un indice d'erreur statistique et d'un facteur de normalisation pour qu'il soit applicable aux différentes variables, en accordant le même poids à tous les débits. Il varie de la valeur optimale « 0 » qui indique une variation résiduelle nulle et donc un modèle de simulation parfait, à des valeurs positives très grandes.

d. Le coefficient de détermination (R^2)

Il permet d'évaluer le degré d'association entre deux variables X et Y et de juger de la qualité de l'ajustement des points par la droite de régression, en expliquant la proportion de la variance de Y par les variations de X. Il peut être estimé par l'équation suivante :

$$R^2 = \frac{\sum_{i=1}^n (Q_{i\text{ obs}} - Q_{i\text{ sim}})^2}{\sum_{i=1}^n (Q_{i\text{ obs}} - Q_{\text{moy}})^2} \quad (3.4)$$

Où Q_{moy} est la moyenne des débits observés. Pour une bonne simulation des débits le coefficient de détermination doit être égal à un.

e. Le critère de Nash

Le critère d'optimisation du modèle est le critère de Nash-Sutcliffe basé sur l'erreur quadratique des débits. Le jeu de paramètres obtenu par l'optimisation de ce critère peut être influencé par de longues périodes présentant de faibles différences entre les séries observées et simulées (Nash et Sutcliffe, 1970).

$$\text{Nash}(Q) = 100 \cdot \left[1 - \frac{\sum_i (Q_{i\text{ obs}} - Q_{i\text{ sim}})^2}{\sum_i (Q_{i\text{ obs}} - Q_{i\text{ sim}})^2} \right] \quad (3.5)$$

Le modèle est parfait pour NASH=100. Il est supposé être acceptable pour NASH>70. Ce critère présente l'inconvénient de donner beaucoup d'importance aux forts débits. Par conséquent, si l'on ne s'intéresse pas spécifiquement aux crues on applique le critère de Nash appliqué aux racines carrées des débits. (Eckert,2002)

f. Percent Bias (PBIAS)

Mesure la tendance moyenne des valeurs simulées (plus grande ou plus petite) par rapport aux valeurs observées, donc estime la sous/surestimation de la simulation. Sa valeur optimale est de « 0 ». Des valeurs positives de ce critère indiquent qu'il y a sous-estimation du biais et des valeurs négatives indiquent qu'il y a surestimation du biais.

$$PBIAS = \frac{\sum_{i=1}^n (Q_{i\text{ sim}} - Q_{i\text{ obs}})}{\sum_{i=1}^n Q_{i\text{ obs}}} \quad (3.6)$$

En fonction des valeurs de différents critères, on peut définir des performances des modèles pluie-débit aux différents pas de temps dans le tableau suivant (Moriassi et al, 2015) :

3.6.5. Etapes d'élaboration d'un modèle

L'élaboration d'un modèle hydrologique comporte certaines étapes essentielles (Boudahraa, 2007) :

- **Définir le modèle et ses objectifs** en termes de précision (Ambroise, 1998). Cette définition permettra de déterminer les choix à faire par la suite et donc de préciser les chemins à suivre dans l'élaboration des étapes suivantes.
- **Identifier le modèle**, pour pouvoir définir et caractériser le système, ses frontières et sa structure ; définir l'événement, les variables et les paramètres ; et émettre les hypothèses et le choix des échelles de temps et d'espace caractéristiques des processus de base.
- **Elaborer l'algorithme** à incorporer dans un outil informatique capable de procéder aux instructions. Cet algorithme doit être suivi de la vérification du logiciel lui-même (De Marsily et Ambroise, 1998).
- **Caler le modèle**, par estimation des paramètres du modèle, non mesurables, selon des critères déductifs ou des observations des entrées et des sorties. Il existe deux façons d'estimer les paramètres à partir d'un échantillon : la méthode passive (on prend tous les couples entrées-sorties disponibles, et par des méthodes statistiques, on évalue les coefficients des équations) et la méthode active (en donnant un poids spécifique plus grand à certaines observations ou groupe d'observations). Le calage peut être manuel ou automatique.
- **Evaluer le modèle**, en comparant la réponse simulée par le modèle à une sollicitation expérimentale et la réponse réelle, à partir de données différentes de celle utilisées pour le calage.
- **Définir le domaine de validité** du modèle élaboré (Ambroise, 1998). La formulation mathématique du modèle est basée sur des équations simplifiées, selon certaines hypothèses concernant les conditions initiales et aux limites.

4. Présentation de quelques outils

4.1. Le logiciel GARDÉNIA

(Modèle Global À Réservoirs pour la simulation de Débits et de Niveaux Aquifères) permet de calculer, à partir de la séquence des données météorologiques (précipitations, évapotranspiration potentielle) sur son bassin d'alimentation :

- le débit à l'exutoire d'un cours d'eau (ou d'une source)
- le niveau piézométrique en un point de la nappe libre sous-jacente

4.2. EMILE

Les stations hydrométriques et les postes pluviométriques en pouvant établir un lien dynamique entre eux autorisent le calcul d'une pluie moyenne sur un bassin versant. Les données traitées peuvent être acquises directement (ETP), ou importées notamment des logiciels de gestion de base de données ORSTOM, HYDRON et PLUVIOM. EMILE permet l'accès à plusieurs - 274 - types de sessions, et donc à différentes utilisations des algorithmes:

calages manuels, calages automatiques, simulations et générations de débits. Les calculs effectués automatiquement utilisent une fonction critère et une méthode d'optimisation numérique.

4.3. Le SIG

Les données extraites de SIG et introduites dans les modèles de simulation (GEO-RAS). Ils ont l'avantage :

- support de l'information géographique
- outil de modélisation et calcul performant
- outils de gestion des données géographique et thématique

5. Applications hydrologiques de modèles pluie-débit

5.1 Le Soil and Water Assessment Tool (SWAT)

Le SWAT est un modèle conceptuel semi- distribué qui fonctionne en continu sur un pas de temps quotidien (Arnold et al., 1998b). Il s'agit d'un outil complet qui permet de prévoir, sur de longues périodes, l'impact des pratiques de gestion des terres sur l'eau, les sédiments et les produits chimiques agricoles pour de grands bassins versants complexes dont les sols, l'utilisation des terres et les pratiques de gestion varient (Neitsch , 2005). SWAT a été développé pour simuler les principaux processus du cycle hydrologique et leurs interactions de manière aussi simple et réaliste que possible, et pour utiliser des données d'entrée facilement disponibles pour les bassins versants à grande échelle, afin de pouvoir être utilisé dans la planification de routine et la prise de décision (Ogden et al., 2001). L'un des principaux avantages de SWAT est qu'il est efficace sur le plan informatique, même pour les plus grands bassins versants, ce qui le rend utile pour les gestionnaires des terres et des ressources en eau. Le modèle a été conçu pour la prédiction des rendements à long terme plutôt que pour des événements de crue uniques (Arnold et al., 1998b).

5.2. Les modèles ' boîte noire '

- **L'hydrogramme unitaire :** La théorie de l'hydrogramme unitaire a été élaborée par L.K. Sherman (1932) (Roche, 1971, Singh, 1995). Elle stipule que le bassin versant est modélisé comme un système soumis à une entrée $E(t)$ (pluie) qui produit l'écoulement et délivrant une sortie $S(t)$ (débit). L'application de la théorie de l'hydrogramme unitaire, nécessite de résoudre le problème de déconvolution, et qui consiste justement à déterminer une fonction de transfert à partir d'un historique assez représentatif des entrées et des sorties. La théorie de l'hydrogramme unitaire reste très limitée aux bassins de faible surface, pour des grandes étendues, cette théorie est impossible à appliquer pour plusieurs raisons, en particulier les pluies et les caractéristiques pédologiques ne sont pas homogènes.

- **Le modèle DPFT (Différence Première de la Fonction de Transfert) :** La méthode DPFT a été développée à l'EDF, et présentée pour la première fois par Duband (1978). Cette méthode est une extension de la méthode classique de l'hydrogramme unitaire, et par conséquent décompose le processus global pluie-débit en modèles distincts par leurs structures, un modèle de production transforme la pluie brute en pluie efficace ruisselée, et une fonction de transfert permettant le passage de la pluie efficace en débit de ruissellement à

l'exutoire du bassin. (Chambaz, 1990, Dechemi & Chambaz, 1994). Deux principes de base s'imposent à cette méthode :

- Le premier est que les débits sont exprimées en différences premières sur un pas de temps reliant le présent et le futur ;
 - Le deuxième est que les pluies efficaces sont considérées comme inconnues du problème.
-
- **Les modèles ARMAX** (Auto-Regressive Moving Average with eXogenous inputs) Les modèles ARMAX dérivent des modèles ARMA (Auto-Regressive Moving Average) initialement développés par Box et Jenkins (1976), ont été utilisés dès leur première conception pour la simulation des débits. Les modèles ARMAX, ont été utilisés dans le contexte pluie-débit, à différents pas de temps, journalier (Weeks & Boughton, 1987), et horaire (Rodriguez, 1989) dans le cadre de la prévision des crues.
 - **Les réseaux de neurones artificiels** : Relevant du domaine de l'intelligence artificielle, les réseaux de neurones sont une métaphore des structures cérébrales, des assemblages de constituants élémentaires appelés neurones fortement connectés construits par couches, qui réalisent chacun une fonction généralement non-linéaire, pour donnée une sortie unique à partir d'un mécanisme d'apprentissage effectué pendant le calage. Les réseaux de neurones n'ont été appliqués qu'après des décennies de recherches, et sont pourvus actuellement de fondements mathématiques très sophistiqués ce qui explique leurs utilisations dans de nombreux domaines scientifiques. L'étude de la relation pluie-débit par les réseaux de neurones a été traitée par plusieurs hydrologues (French et al. 1992, Hsu et al., 1995, Shamseldin 1996, Dimopoulos et al., 1996, Minns & Hall 1996, Coulibaly et al., 1999, Anctil et al., 2004).
 - **La logique floue** : La logique floue introduite par Zadeh (1965), est un langage naturel issu de la théorie des ensembles, programmée afin de pallier aux insuffisances de la logique classique. Celle-ci n'utilisant que deux états logiques Vrai et Faux, en langage machine cela donnera les deux valeurs 0 et 1. A partir des règles floues, le processus à modéliser est divisé en un ensemble de processus plus simples (sous ensembles flous), et en fonction des degrés d'appartenance (degrés de vérité), on engendre des surfaces d'inférences, les valeurs sont obtenues par la projection du centre de gravité de ces surfaces. La logique floue, par son pouvoir d'extrapolation non-linéaire a été exploitée dans la modélisation pluie-débit (Özelkan & Duckstein, 2001, Dechemi et al., 2003, Benkaci & Dechemi, 2004), et s'est vue une utilisation renommée en hydrologie.
 - **Le Modèle GR** : Les modèles de simulation du Génie Rural fonctionnent aux pas de temps annuel, mensuel et journalier.

A. Le modèle pluie-débit annuel GR1A

Le modèle GR1A « modèle du Génie Rural à 1 paramètre Annuel » est un modèle pluie-débit global. Son développement a été initié au Cemagref à la fin des années 1990. La version que nous avons utilisée, est celle proposée par Mouelhi (2003) et Mouelhi et al. (a, b 2006)

Modélisation hydrologique

La structure du modèle est très simple puisqu'elle se résume à une simple équation, le débit Q_k de l'année k étant proportionnelle à la pluie P_k de la même année, avec un coefficient d'écoulement dépendant de P_k , de la pluie P_{k-1} de l'année $k-1$ et de l'évapotranspiration potentielle annuelle moyenne E . Le modèle s'écrit :

$$Q_k = \left\{ 1 - \frac{1}{\left\{ 1 + \left(\frac{0.7P_k + 0.3P_{k-1}}{X.E_k} \right)^2 \right\}^{0.5}} \right\} \quad (3.7)$$

Où :

- Q_k est le débit simulé de l'année k
- P_k est la pluie observée de l'année k
- P_{k-1} est la pluie observée de l'année $k-1$
- E_k est l'évapotranspiration potentielle de l'année k
- X est le paramètre du modèle à optimiser

B. Modèle pluie-débit mensuel GR2M :

Le modèle GR2M (modèle Mensuel du Génie Rural à 2 paramètres) est un modèle pluie-débit global.

La fonction de production du modèle repose sur un réservoir de suivi d'humidité du sol. Une partie P_s de la pluie P_k va être ajoutée au contenu S_k dans le réservoir en début de pas de temps. Le paramètre X_1 , capacité du réservoir, est positif et exprimé en mm. La pluie en excès, P_1 , est donnée par :

$$P_1 = P - P_s \quad (3.8)$$

et le contenu du réservoir est actualisé :

$$S^k = S_k + P_s \quad (3.9)$$



Mise
en
application

Exercice

Afin d’initier les étudiants à la modélisation hydrologique, des mini projets ont été réalisés. Dans ce qui suit, un exemple d’application du modèles GR1A et GR2M au niveau du bassin versant d’oued Sebdou.

I. Les données

Afin de prédire le débit pour une année donnée, on utilise à l'entrée du modèle des valeurs des pluies observées (mm), des ETP calculées par la méthode Thornthwaite (mm) et des débits observés (exprimés en lames d'eau écoulées).

Tableau 3.1: Caractéristiques des stations hydroclimatiques utilisées

	code	station	X	Y	Altitude (m)	Période d'observation
			(km)	(km)		
Stations hydrométriques	160401	Sebdou	131.25	156.57	875	1971-2017
	160402	Beni Bahdel	119	163.3	665	1971-2017
	160409	Zahra	114.65	161.5	660	1971-2017
Stations pluviométriques	160403	Beni Bahdel	114.98	164.60	666	1975-2013
	160401	Sebdou	131.25	158.55	875	1975-2013
	160406	Khemis,O Moussa	109.30	157.00	870	1975-2013
	160407	Sidi Medjahed	102.90	173.10	360	1975-2013
Stations climatiques	160507	Zenata	115.0	164.6	115.0	1981-2016

Une critique rigoureuse et un traitement laborieux ont été effectués avant l’utilisation des séries d’observation (Comblement des lacunes, étude d’homogénéité, ...).

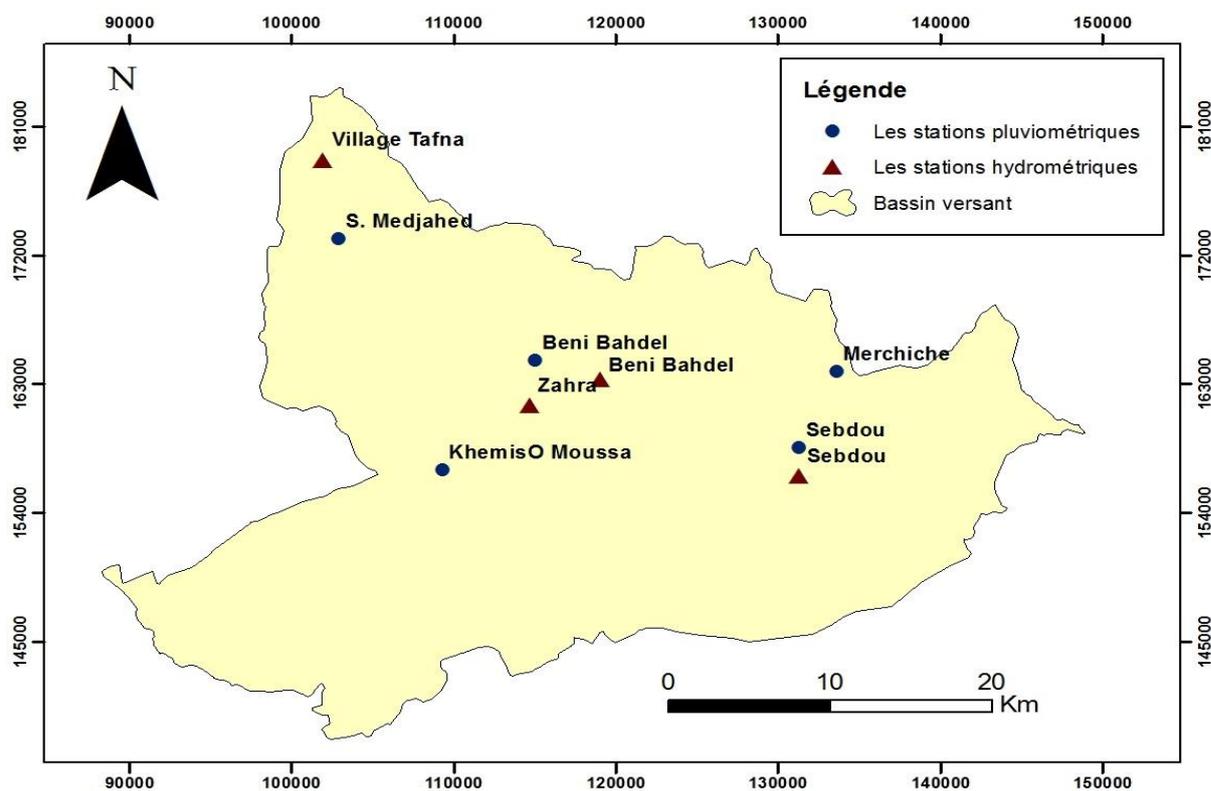


Figure 3.5 : répartition des stations utilisées

Tableau 3.2: les données hydroclimatiques utilisées

Année	Q (mm)	P(mm)	Q(m3/s)	année	Q (mm)	P(mm)	Q(m3/s)
1981 - 1982	10,62	339,90	0,09	1999 - 2000	15,39	332,80	0,13
1982 - 1983	17,73	279,40	0,14	2000 - 2001	146,21	417,80	1,18
1983 - 1984	6,16	247,60	0,05	2001 - 2002	25,04	347,90	0,20
1984 - 1985	11,64	329,10	0,09	2002 - 2003	59,07	463,10	0,49
1985 - 1986	48,58	390,60	0,41	2003 - 2004	19,81	338,70	0,16
1986 - 1987	105,91	430,70	0,88	2004 - 2005	11,48	366,60	0,09
1987 - 1988	8,32	231,50	0,07	2005 - 2006	38,58	394,50	0,31
1988 - 1989	13,96	337,80	0,11	2006 - 2007	5,97	327,00	0,04
1989 - 1990	6,78	265,70	0,05	2007 - 2008	17,10	272,40	0,14
1990 - 1991	74,13	451,80	0,60	2008 - 2009	162,28	658,90	1,32
1991 - 1992	27,28	338,45	0,22	2009 - 2010	34,59	389,20	0,28
1992 - 1993	8,47	225,10	0,07	2010 - 2011	3,92	360,90	0,03
1993 - 1994	10,33	286,90	0,09	2011 - 2012	9,23	293,70	0,08
1994 - 1995	35,36	351,10	0,28	2012 - 2013	89,41	548,30	0,73
1995 - 1996	85,69	694,00	0,70	2013 - 2014	60,31	593,20	0,51
1996 - 1997	5,51	518,40	0,04	2014 - 2015	45,79	437,80	0,38
1997 - 1998	11,23	342,80	0,09	2015 - 2016	33,74	376,00	0,27
1998 - 1999	15,70	247,80	0,13				

II. Etapes de réalisation du modèle

1. Calcul de l'évapotranspiration potentielle (ETP)

C'est la quantité d'eau susceptible d'être évaporée par une surface d'eau libre ou par un couvert végétal dont l'alimentation en eau n'est pas le facteur limitant. La valeur de l'évapotranspiration potentielle est fonction de l'énergie disponible (Margat, 1997).

a. Méthode de Thornthwaite

Pour estimer l'ETP, en 1948 Thornthwaite a établi une corrélation entre la température moyenne mensuelle et l'évapotranspiration potentielle mensuelle par la formule suivante :

$$\text{ETP non corrigé} = 1.6 \left(\frac{10T}{I} \right)^a \quad (3.10)$$

- $I = \sum_1^{12} i \quad (3.11)$

- $i = \left(\frac{T}{5} \right)^{1.514} \quad (3.12)$

- $a = \left(\frac{1.6 \cdot I}{100} \right) + 0.5 \quad (3.13)$

- $k = 0.49 + 1.8 \left(\frac{I}{100} \right) - 0.77 \left(\frac{I}{100} \right)^2 + 0.67 \left(\frac{I}{100} \right)^3 \quad (3.14)$

$$\text{ETP corrigé} = 1.6 \left(\frac{10T}{I} \right)^a \cdot k \quad (3.15)$$

Avec :

ETP : Evapotranspiration potentielle (mensuelle) (mm),

T : Température moyenne mensuelle du mois considéré (°C),

I : Somme des 12 indices thermiques mensuels,

K : Coefficient de correction liée à la latitude (durée maxi de l'ensoleillement).

Remarque : Pour le calcul de l'ETP par la méthode de Thornthwaite, nous avons utilisé la feuille de calcul téléchargée via le lien suivant : <https://webgr.inrae.fr/modeles/modele-devapotranspiration/>

Tableau 3.3: Valeurs d'ETP par la méthode de Thorntwaite (1981-2016)

Années	ETP (mm)	Années	ETP (mm)	Années	ETP (mm)	Années	ETP (mm)
1981	1264.01	1990	1477.34	1999	1464.98	2008	1466.00
1982	1410.78	1991	1646.08	2000	1430.68	2009	1417.93
1983	1308.33	1992	1439.42	2001	1393.36	2010	1344.29
1984	1362.19	1993	1437.34	2002	1288.15	2011	1456.16
1985	1380.05	1994	1521.07	2003	1522.34	2012	1660.14
1986	1481.05	1995	1235.84	2004	1517.06	2013	1398.22
1987	1314.83	1996	1267.89	2005	1570.11	2014	1327.53
1988	1358.19	1997	1187.44	2006	1410.14	2015	1525.51
1989	1399.11	1998	1414.03	2007	1408.32	2016	1331.04

2. Evapotranspiration réelle (ETR)

Méthode de Turc

La formule de Turc pour le calcul de l'ETR est la suivante:

$$ETR = \frac{P}{\sqrt{0.9 + \frac{P^2}{L^2}}} \quad (3.16)$$

- ETR : Evapotranspiration réelle (mm/an)
- $L = 300 + 25T + 0,05T^3$
- P : Précipitations moyennes annuelles (mm)
- T : Température moyenne annuelle (°C)

Tableau 3.4: Calcul de l'ETR par la méthode de Turc

P (mm)	T (°C)	L	ETR (mm)	Écoulement (mm)
372.47	17.7	1019.76	368.78	4.1

3. Calcul du ruissellement « R »

Le ruissellement de surface est la partie de l'eau de pluie qui ne revient pas par l'évaporation ou l'atmosphère qui n'est pas absorbée par le sol et ne s'accumule pas, mais il traverse la surface du sol par de petits canaux jusqu'à ce qu'il atteigne le réseau de drainage naturel du bassin versant. Le nombre de couches d'écoulement d'eau dépend de l'intensité et de la hauteur de la couche, séquence des précipitations, état d'humidité et structure des précipitations ; le niveau de surface, le couvert végétal, la pente et la rugosité du sol avant la

pluie de terrain. L'effet de tous ces facteurs rend le ruissellement difficile et son estimation précise.

Il peut être estimé par la méthode de Tixeront – Berkaloff selon la formule :

$$R = \frac{P^3}{3ETP^2} \quad (3.17)$$

avec :

R : Ruissellement (m)

P : Précipitations (m)

ETP : Evapotranspiration potentielle (m)

Tableau 3.5 : Valeur du ruissellement

Station	P (m)	T (°C)	ETP (m)	R (m)
Sebdou	0,372	17,7	1.41	0.034

Le ruissellement calculé à partir de la méthode de Tixeront–Berkaloff est de 34 mm.

4. Calcul du bilan hydrologique par la méthode de Thorthwaite

La présence de données des précipitations et des évapotranspirations potentielles nous permettent également d'établir le bilan hydrologique comme suit:

Si l'évapotranspiration réelle est liée à la réserve en eau disponible nous considérons deux possibilités :

Possibilité	ETR = ETP	ETR = ETP	ETR = P
Condition	P > ETP	P < ETP et existence de réserve	P < ETP et pas de réserve

Nous considérons une réserve utile de 100mm (Baba Hamed, 1991).

Tableau 3.6: Eléments du bilan hydrologique calculés pour la station de Sebdou (1971-2016)

Mois	S	O	N	D	J	F	M	A	M	J	Jt	A	Annuel
P	20.53	29.88	45.34	45.9	49.01	48.54	43.4	34.69	32.02	10.45	4.82	7.92	372.47
ETP	2329.4	1085.4	342.7	143.8	92.9	125.4	293.3	509.4	1152.9	2461.1	4169.3	4240.1	1412.1
	-2308	-1055	-297	-97.9	-43	-76	-249	-474	-1120	-2450	-4164	-4232	-----
RU	0	0	1.5	18.4	50.6	72.4	74.7	50.9	0	0	0	0	-----
ETR	20.53	29.88	342.7	143.8	92.9	125.4	293.3	509.4	32.02	10.45	4.82	7.92	134.43
DA	2308.9	1055.5	0	0	0	0	0	0	1120.9	2450.7	4164.5	4232.2	1277.7

P : précipitations (mm), ETP : évapotranspiration potentielle (mm), U: variation de la réserve, RU: réserve utile (mm)
 ETR : évapotranspiration réelle (mm), D.A: déficit agricole (mm).

D'après le tableau 3.6, on remarque que :

Le déficit agricole est nulle à partir du mois de Novembre jusqu'à Avril ce qui indique que le bilan hydrologique est équilibré. Pour les autres mois de l'année, le déficit agricole varie entre 1055.5 mm et 4232.2 mm indiquent la rareté des ressources en eau.

Ce déficit hydrique marqué, dont l'origine se trouve essentiellement dans la faiblesse des précipitations sur une période prolongée par rapport à la moyenne des apports observés sur cette période. Ce manque de pluie a une incidence directe sur la végétation cultivée : on parle alors de sécheresse agricole ou édaphique (liée à la réserve en eau du sol). La contrainte hydrique est souvent associée à des contraintes thermiques (hautes températures). On déduit que l'existe une incapacité de relayer suffisamment le défaut de ressource édaphique et pluviométrique par une irrigation. On parle alors de La sécheresse qui peut donc concerner l'agriculture directement par un déficit pluviométrique en cours de saison de végétation (en gros, printemps et été) ou par une reconstitution déficiente des réserves hydrographiques suite à un déficit pluviométrique d'automne ou d'hiver. On peut qualifier la première situation de sécheresse agricole ou édaphique et la deuxième de sécheresse hydrologique. Ces résultats peuvent être un outil d'aide à la décision pour les gestionnaires du secteur agricole de la région.

5. Elaboration du modèle GR1A :

Remarque : La feuille de calcul disponible via le lien suivant a été utilisée : <https://webgr.inrae.fr/modeles/annuel-gr1a/>

✚ Pour le calage, nous avons procédé à des changements sur le paramètre X , en commençant par la valeur 0.13 et la faisant accroître successivement avec un intervalle de 0.01 jusqu'à la valeur 3.5, et d'un critère de Nash supérieur ou égale à 70% , et un coefficient de détermination R^2 entre les débits mesurés et ceux calculés proche de 1. Nous avons obtenue le meilleur calage présenté dans le tableau suivant :

Tableau 3.7: Résultat du calage du modèle GR1A au bassin versant d'Oued Sebdou (1981-2005)

Critères	X	Nash (%)	Bilan (%)	Coefficient de détermination (R^2)
Sebdou	1.03	81.2	97.3	0.8078

A partir des valeurs des coefficients de détermination, le calage parait correct. On remarque que la valeur du Nash est de 80.78%, les débits observés et simulés sont en parfaite superpositions. Le bilan hydrologique a été simulé avec un pourcentage très satisfaisant 97.3% ce qui nous amène à conclure qu'il est équilibré. Le coefficient de correction de l'évapotranspiration X_1 est supérieur à 1 ce qui nous laisse penser que l'évapotranspiration est importante. Il traduit aussi les échanges avec des nappes profondes d'où une infiltration souterraine vers l'alimentation des nappes tout au long de l'oued en raison du caractère karstique du bassin étudié.

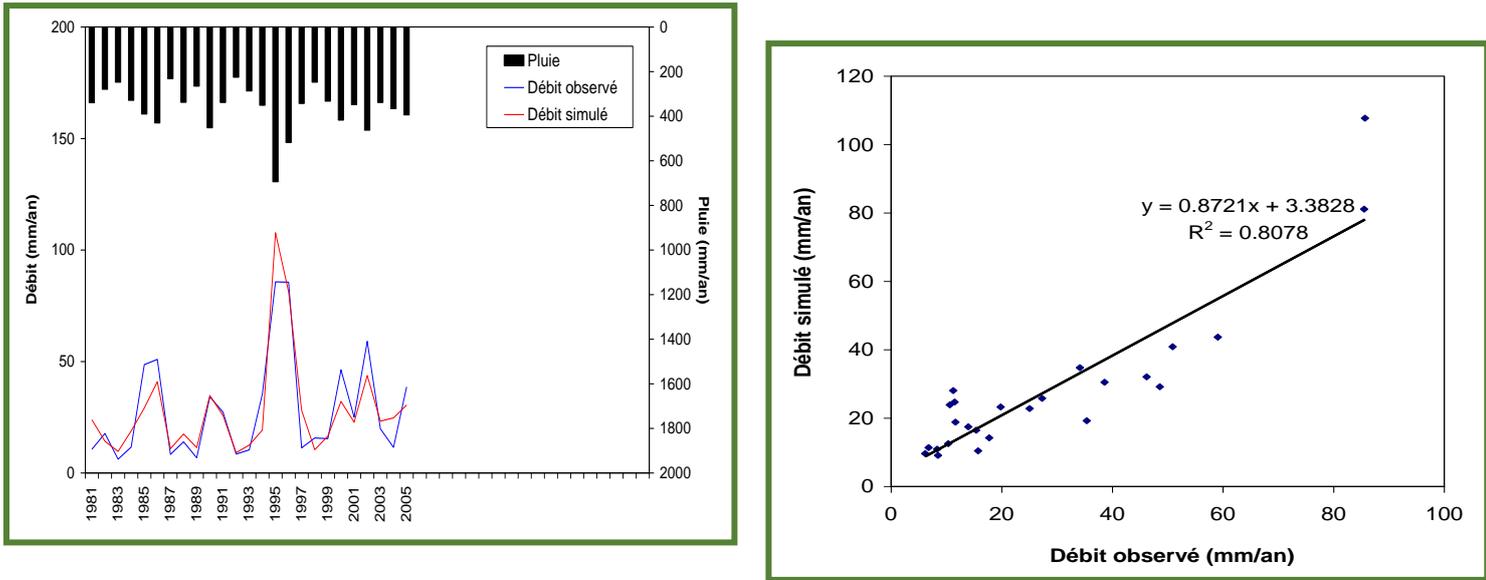


Figure 3.6 : Résultats du calage obtenus par le modèle GR1A, Corrélation entre les débits observés (Qobs) et les débits simulés (Qsim), Visualisation de la qualité du calage au bassin versant d’Oued Sebdou (1981-2005)

✚ Pour la validation du modèle de nouvelles données des précipitations et d’évapotranspiration correspondant à la période (2006-2015) et qui n’ont pas été utilisées lors du calage sont introduites. Le calcul est lancé en prenant pour le paramètre X la valeur trouvée lors du calage. Les valeurs trouvées des débits sont alors comparées aux valeurs observées par corrélation linéaire simple (Figure III. 2). Les résultats sont présentés dans ce tableau suivant :

Tableau 3.8: Résultats de la validation du modèle GR1A au bassin versant d’Oued Sebdou (2006-2015)

Critères	Nash (Q) %	Coefficient de corrélation (R)
Sebdou	83.3	0.85

A partir des résultats obtenus la modélisation pluie-débit élaborée par utilisation du modèle GR1A donne des résultats satisfaisants. La corrélation entre les débits observés et ceux simulés est significative. La majorité des points s’alignent autour de la droite à l’exception de quelques points. On constate aussi un synchronisme entre les courbes des lames d’eau écoulées calculées et celles mesurées qui parfois sont complètement superposées.

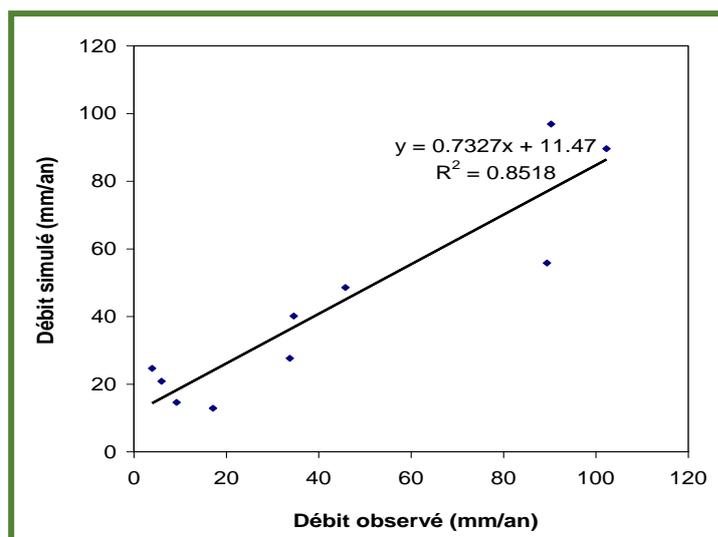


Figure 3.7 : Validation du modèle GR1A au bassin versant d'Oued Sebdou (2006-2015)

6. Elaboration du modèle mensuel GR2M

Remarque : La feuille de calcul disponible via le lien suivant a été utilisée :

<https://webgr.inrae.fr/modeles/mensuel-gr2m/>

Le modèle GR2M (modèle du Génie Rural à 2 paramètres Mensuel) est un modèle pluie-débit global à deux paramètres optimisables : X_1 : capacité du réservoir de production (mm), X_2 : coefficient d'échanges souterrains (mm). Il fonctionne autour de deux réservoirs, un de production (ou réservoir sol) gère, en fonction de son taux de remplissage, la part de pluie qui se transforme en pluie nette et un de routage, associé à un hydrogramme unitaire, permet d'étaler le temps de cette pluie nette pour obtenir le débit de sortie et sur lesquels les ajustements et interceptions se font différemment sur les entrées. Il s'est distingué comme le modèle fournissant globalement les meilleurs résultats de modélisation.

Pour prédire le débit à un mois donné nous avons utilisé à l'entrée du modèle des valeurs des pluies mensuelles observées (mm), des ETP calculées par la méthode de Thornthwaite (mm) et des débits mensuels observés exprimés en lames d'eau écoulées (mm).

✚ Pour le calage du modèle, nous avons procédé manuellement aux changements des valeurs des paramètres X_1 et X_2 plusieurs fois jusqu'à l'obtention des valeurs optimales du coefficient du critère de qualité de Nash Sutcliffe et du coefficient de détermination R^2 de la corrélation entre les débits calculés et ceux observés.

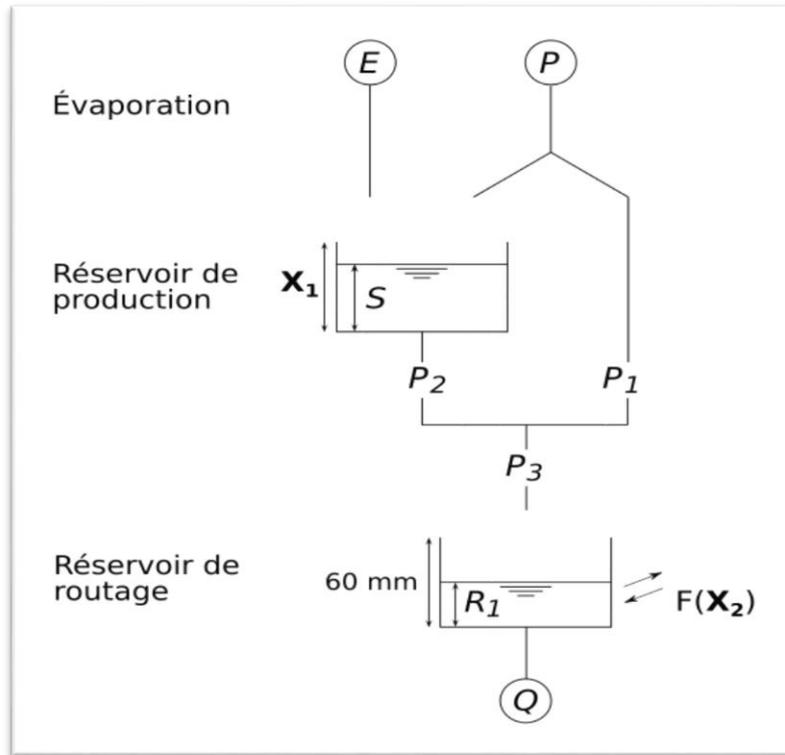


Figure 3.8: Organigramme du modèle GR2M

D'après les résultats, la valeur du paramètre X_1 (capacité du réservoir de production) respecte les bornes de l'intervalle de confiance, de même pour la valeur du paramètre X_2 (coefficient d'échanges souterrains) qui est positif, cela indique que le modèle simule un apport d'eau provenant des bassins voisins. En effet, on peut considérer qu'avec l'arrivée de la pluie au sol, les échanges avec les nappes profondes se déclenchent, d'où une infiltration souterraine vers l'alimentation des nappes tout au long de l'oued, accentuée par le caractère karstique du bassin. Les valeurs du critère de Nash varient est 90.1% et les valeurs du coefficient de détermination $R^2 = 0.83$. Le bilan assez équilibré. Les bonnes superpositions des deux courbes celle des débits simulés et calculés permettent de dire que les modèles est bien calé.

Tableau 3.9 : Résultat du calage du modèle GR2M au bassin versant d'Oued Sebdou (1981-2005)

Période de calage (01/1981-12/2010)				
Paramètres		Critères de qualité et d'optimisation		
X_1	X_2	Nash (Q) %	Bilan %	Coefficient de corrélation R
221.4	0.69	90.1	99.7	0.83

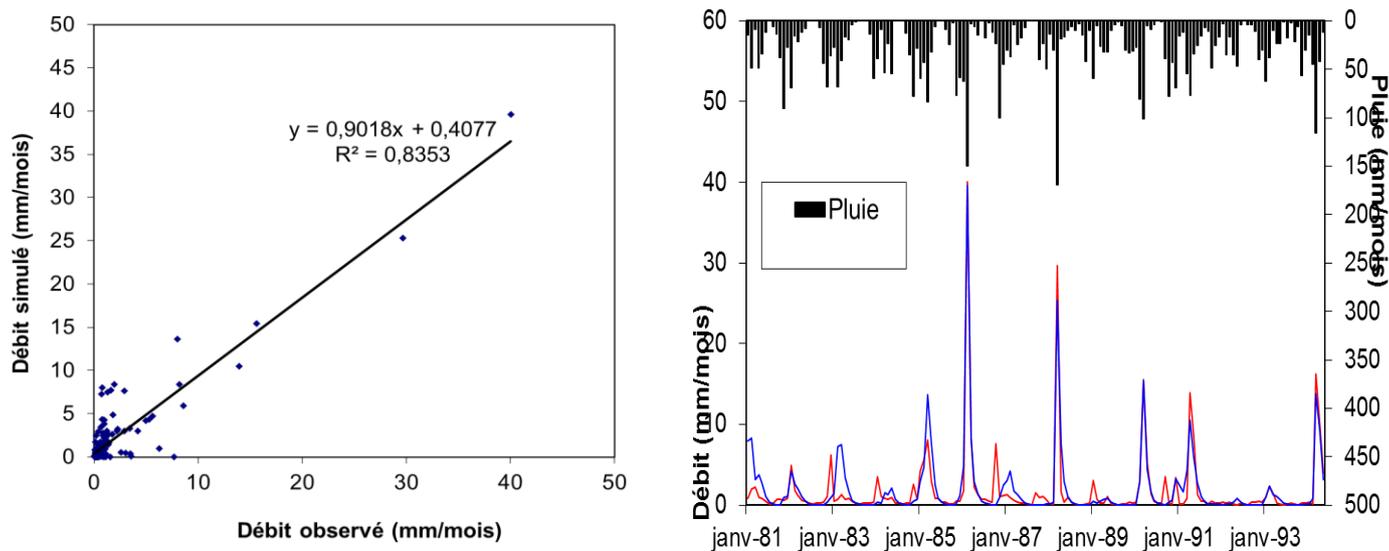


Figure 3.9: Résultats du calage obtenus par le modèle GR2M, Corrélation entre les débits observés (Qobs) et les débits simulés (Qsim), Visualisation de la qualité du calage au bassin versant d’Oued Seb dou (1981-1993)

La validation des débits simulés en fonction des débits observés donne des valeurs du coefficient de détermination $R^2=0.81$. La valeur de Nash est acceptable, par conséquent nous pouvons dire que la modélisation pluie – débit par l’utilisation du modèle GR2M donne des résultats acceptables et très encourageants.

Tableau 3.10: Représentation des résultats de la validation obtenue par utilisation de GR2M sur le bassin versant de Seb dou

Période de validation (01/2011-12/2015)		
Critère de qualité et d’optimisation		
Nash (Q) %	Bilan %	Coefficient de corrélation R
83	105.7	0.81

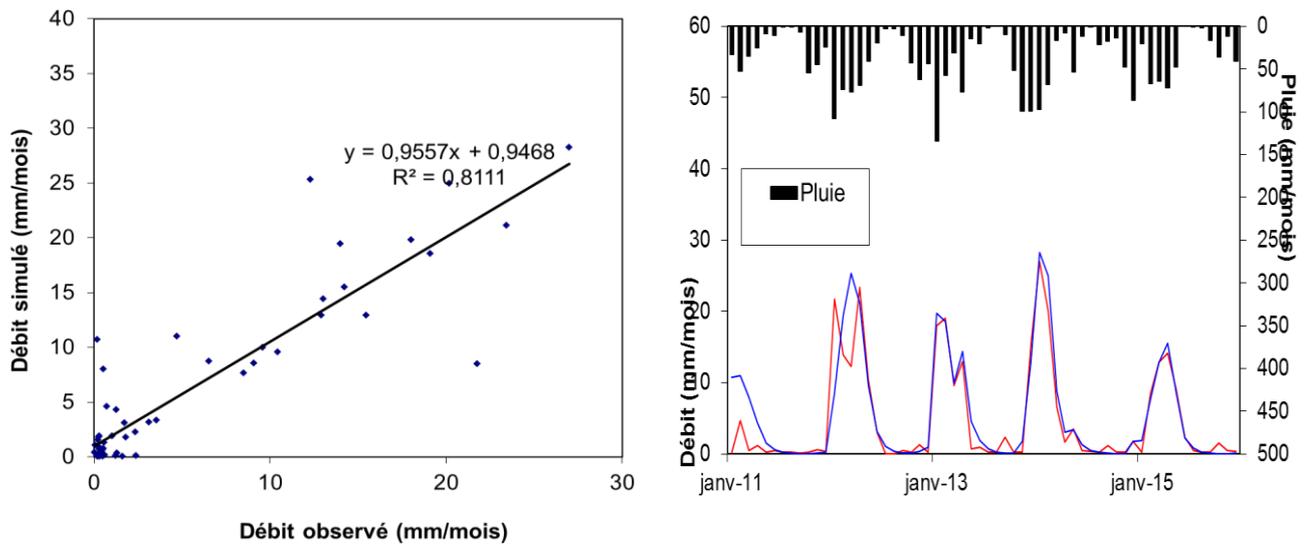


Figure 3.10 : Validation du modèle GR2M au bassin versant d'Oued Sebdu (2011-2015)

Sujets proposés

Examen (2019/2020)

Exercice 1

Soit une série de pluies annuelles (Tableau 1) ayant une moyenne de 509.19 mm et un écart type de 146.19 mm.

1. En supposant que la série d'observations s'ajuste à une loi normale. Calculer la pluie décennale et centennale.
2. Vérifier la normalité de cette série en appliquant le test de khi carré au seuil de 5%.

Tableau 1

Classe	Effectif n_i
[787.3-526.2]	5
]526.2-486.3]	5
]486.3-299]	5

Exercice 2

Soient deux séries dont les caractéristiques statistiques sont représentées dans le tableau 2.

1. Quelles sont les conditions à vérifier pour étudier la corrélation entre la série 1 par rapport à la série 2.
2. Calculer le coefficient de corrélation. Que déduisez-vous ?
3. Déterminer la droite de régression.

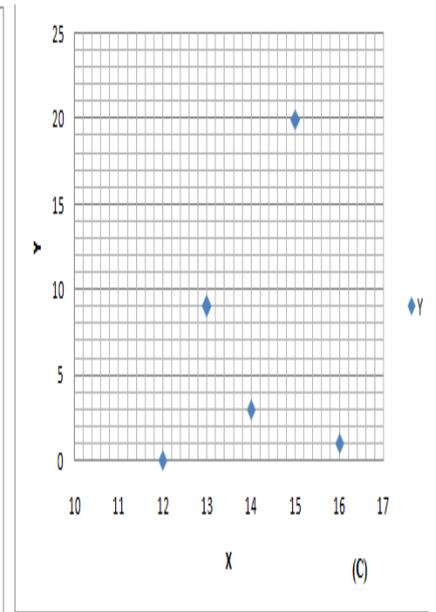
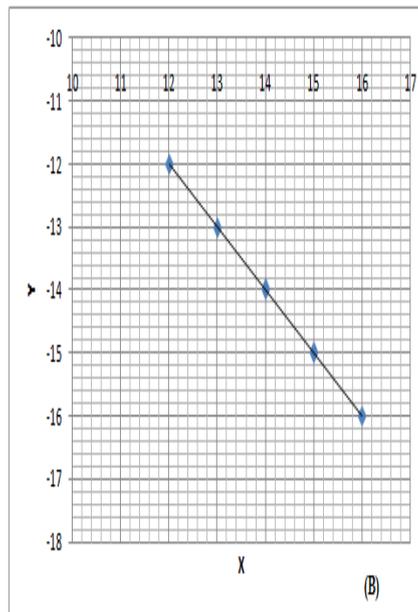
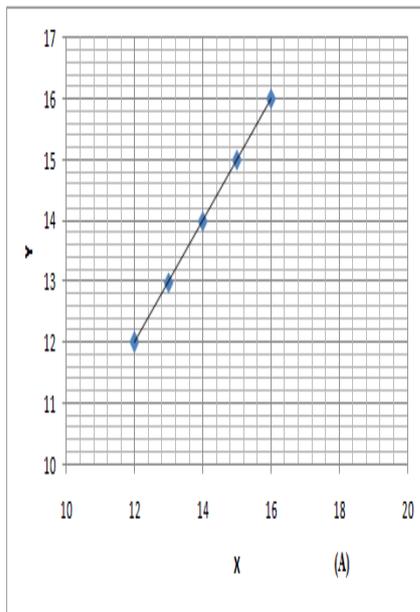
Tableau 2

N°	Série 1	Série 2
1	345.3	252.6
2	534.7	326
3	353.6	383
4	467.9	368.5
5	336.9	290
6	551.2	367.4
7	232.9	235.5
8	465.1	366.7
9	391.7	329.6
10	236.2	235.7
Moyenne	391.6	315.5
Ecart type	106.3	55.1

Examen (2020/2021)

Questions de cours

- Donner la définition de la corrélation entre deux variables.
- Pourquoi calcule-t-on la corrélation entre deux variables ?
- Quelles sont les valeurs du coefficient de corrélation pour les cas (A), (B) et (C) ?



Exercice 1

Soit une série de pluies annuelles (Tableau 1).

1. En supposant que la série d'observations s'ajuste à la loi de Gumbel. Trouver la droite d'ajustement.
2. Calculer la pluie décennale et la pluie centennale.

Tableau 1

Année	Pluie annuelle (mm)	Année	Pluie annuelle (mm)
1974	43.32	1982	512
1975	206.4	1983	42.6
1976	324	1984	265.4
1977	135	1985	254.4
1978	154.2	1986	270
1979	0.004	1987	323.6
1980	313.4	1988	512
1981	243.8	1989	253.4
moyenne		240.8	
écart-type		147.0	

Exercice 2.

En appliquant le test de Kolmogorov Smirnov au seuil de 5%, Vérifier la normalité de la série représentée dans le tableau 2.

Tableau 2

Année	Station x	Année	Station x
1972	11.1	1979	34.1
1973	10.7	1980	83.0
1974	35.2	1981	20.5
1975	49.7	1982	9.2
1976	27.1	1983	3.3
1977	4.4	1984	5.5
1978	31.1	1985	65.2
Moyenne		27.9	
Ecart-type		24.3	

Examen (2021/2022)

Exercice 1

Soit une série de pluies annuelles dont la moyenne $\bar{x} = 257,2$ mm et l'écart-type $\sigma = 79,60$ mm.

3. En supposant que la série d'observations s'ajuste à la loi Normale, Trouver la droite d'ajustement.
4. Calculer la pluie décennale et la pluie centennale.

La série de pluies annuelles a été classée par ordre décroissant et les valeurs ont été répertoriées en 4 classes (Tableau 1).

Tableau 1 : Répartition des pluies annuelles en classes

Numéro de la classe	Bornes de la classe	Effectif observé n_i
1	[321,1- 466,4 [5
2	[251,9 - 321,1[5
3	[225,0- 251,9 [5
4	[110,8- 225,0 [6

Appliquer le test de **KHI Carré** en répondant aux questions suivantes :

1. Calculer les variables centrées réduites U_i et U_{i+1} correspondant à la borne inférieure et la borne supérieure de la classe i respectivement.
2. En utilisant la table de Gauss, Calculer les fréquences théoriques $F(U_i)$ et $F(U_{i+1})$.
3. Calculer l'effectif théorique U_i correspondant à chaque classe i .
4. Calculer le **KHI Carré**.
5. Vérifier la normalité de la série au seuil de 5%.

Exercice 2

Soient deux séries dont les caractéristiques statistiques sont représentées dans le tableau 2.

4. Quelles sont les conditions à vérifier pour étudier la corrélation entre la série 1 par rapport à la série 2.
5. Appliquer le test de Wilcoxon pour vérifier l'homogénéité de la série 1 par rapport à la série 2.

Tableau 2

Année	Série 1	Série 2
2001	45	40
2002	100	97
2003	310	300
2004	258	250
2005	654	469
2006	27	23
2007	12	5
2008	680	678
2009	99	98
2010	80	79
La moyenne	226.5	222,1
L'écart type	251.4	203,9

Exercice 3

Soit une série de pluies annuelles (Tableau 1).

1. En supposant que la série d'observations s'ajuste à la loi de Galton. Trouver la droite d'ajustement.
2. Calculer la pluie décennale et la pluie centennale.
3. Calculer l'intervalle de confiance de la moyenne de la série étudiée.
4. En appliquant le test de Khi Carré au seuil de 5%, Vérifier la normalité de la série.

Tableau 1

Année	Pluie annuelle (mm)	Année	Pluie annuelle (mm)
1974	43.32	1982	512
1975	206.4	1983	42.6
1976	324	1984	265.4
1977	135	1985	254.4
1978	154.2	1986	270
1979	0.004	1987	323.6
1980	313.4	1988	512
1981	243.8	1989	253.4
	moyenne		240.8
	écart-type		147.0

Examen (2022/2023)

Exercice 1

Soit une série de pluies annuelles ayant une moyenne de 460 mm et un écart type de 126 mm.

5. En supposant que la série d'observations s'ajuste à une loi normale. Calculer la pluie décennale et centennale.
6. Quelle est la probabilité pour que $P > 400$ mm ? Calculer la période de retour.
7. Calculer l'intervalle de confiance de la moyenne et de l'écart type si le nombre d'observation $N = 60$ ans et le seuil de sécurité $\alpha = 5\%$.

Exercice 2

Vérifier l'ajustement de l'échantillon représenté dans le tableau ci-dessous à la loi normale en appliquant le test de Kolmogorov Smirnov au seuil de 5%. Sachant que $\bar{x} = 440,3$ mm et $\sigma = 92$ mm.

Tableau 1

i	X _{classée}	i	X _{classée}
1	319,2	11	423
2	338,6	12	466,9
3	339,3	13	469,6
4	351	14	485,1
5	355,2	15	491,5
6	360,1	16	512,3
7	382,9	17	535,6
8	387,6	18	559,6
9	400,2	19	589,7
10	403,7	20	634,4

Conclusion

L'application de modèles mathématiques aux observations et aux expérimentations hydrologiques sur de petits bassins versants est le complément logique et nécessaire à l'interprétation de leur fonctionnement. On a pris l'habitude de classer les modèles dans des catégories qui, bien que ce classement ne soit pas toujours exclusif, permettent de mieux décrire l'opération réalisée. On dit qu'un modèle est :

- déterministe (par opposition à stochastique), lorsque les variables employées ne sont pas aléatoires et résultent soit de l'observation, soit de grandeurs reconstituées.
- conceptuel (par opposition à empirique), si les algorithmes qui constituent le modèle s'essayent à représenter de façon plus ou moins précises les phénomènes physiques que l'on cherche à décrire.
- global, si la modélisation concerne l'entité géographique complète étudiée, sans modulation selon les particularités locales.
- distribué, par opposition à global, si le modèle résulte de la composition de fonctions ponctuelles ou locales pour l'obtention d'un résultat concernant l'entité géographique totale.

Un modèle hydrologique peut permettre d'estimer les débits en rivière en tout point du bassin versant. La modélisation hydrologique est un moyen pour évaluer et contrôler l'impact des aménagements opérés sur le territoire d'un bassin versant ; notamment les processus rapides d'urbanisation et d'industrialisation ; sur le régime hydrologique.

Ce cours a été rédigé pour permettre aux lecteurs d'explorer le monde attrayant de la modélisation hydrologique, en axant sur les fondements et les principes de base. Dans un contexte où ce genre de documentation est peu abondant, nous espérons que ce modeste travail soit une éventuelle référence.

Références bibliographiques

- Abbott, M. B., Bathurst, J. C., Cunge, J. A., O'Connell, P. E., and Rasmussen, J. (1986). An introduction to the European Hydrological System — Systeme Hydrologique Europeen, —SHEll, 1 : History and philosophy of a physically-based, distributed modelling system. *Journal of Hydrology*, 87(1), 45-59. [https://doi.org/10.1016/0022-1694\(86\)90114-9](https://doi.org/10.1016/0022-1694(86)90114-9)
- Allafta, H., and Opp, C. (2021). GIS-based multi-criteria analysis for flood prone areas mapping in the trans-boundary Shatt Al-Arab basin, Iraq-Iran. *Geomatics, Natural Hazards and Risk*, 12(1), 2087-2116. <https://doi.org/10.1080/19475705.2021.1955755>
- Allen Freeze, R. (1987). Hydrological forecasting, M. G. Anderson and T. P. Burt (eds), Wiley, Chichester, 1985. No. of pages : 616. price £36.50. ISBN 0 471 90614 X. *Earth Surface Processes and Landforms*, 12(2), 214-215. <https://doi.org/10.1002/esp.3290120214>
- Ambroise, B. (1998). Genèse des débits dans les petits bassins versants ruraux en milieu tempéré : 1-processus et facteurs. *Revue des sciences de l'eau/Journal of Water Science*, 11(4), 471-49
- Anctil, F., Michel, C., Perrin, C. and Andréassian, V. (2004) A soil moisture index as an auxiliary ANN input for stream flow forecasting. *Journal of hydrology*, 286; 155-167.
- Arnold, J. G., Srinivasan, R., Muttiah, R. S., and Williams, J. R. (1998b) Large Area Hydrologic Modeling and Assessment Part I : Model Development1. *JAWRA Journal of the American Water Resources Association*, 34(1), 73-89. <https://doi.org/10.1111/j.1752-1688.1998.tb05961>.
- Baba- Hamed, K., & Kerzabi, R.(1991) Contribution à l'Etude hydrogéologique de la haute Tafna en amont du barrage BENI – BAHDEL (MONTES DE TLEMCEEN). Mémoire d'ingénieur d'état. Université de Tlemcen.
- Benkaci, T. (2006) Modélisation pluie-débit mensuelle et journalière par les modèles conceptuels et les systèmes neuro-flous (Application aux bassins algériens), Thèse de doctorat, Institut National d'Agronomie- Algérie
- Benkaci, Ali T., et Dechemi, N. (2004) Modélisation pluie-débit journalière par les modèles conceptuels et 'boit noire', test d'un modèle neuroflou. *Journal des sciences hydrologiques* 4(21).
- Berthier, C.H (2005) Quantification des incertitudes des débits calculés par un modèle pluiedébit empirique. Mémoire de Master 2 en Hydrologie, Hydrogéologie et sols, Université Paris Sud 11.
- Beven, K.J, Warren, R., Zaoui, J. (1980) SHE: towards a methodology for physically based distributed forecasting in hydrology. *Hydrological forecasting (Proceedings of the Oxford Symposium): IAHS – AISH Publ. n° 129*, pp. 133-137.
- Biao, Shi et al (2016) New fuzzy neural network–Markov model and application in mid- to long-term runoff forecast. *Hydrological Sciences Journal*, vol 61, issue 6, 1157-1169 pp.

Références bibliographiques

- Bois, P. (2003) *Hydraulique des écoulements en rivière. Deuxième Partie.* Grenoble.
- Boudahraa, H. (2007) *Modélisation pluie - débit à base géomorphologique en milieu semi-aride rural Tunisien : Association d'approches directe et inverse.* Thèse de Doctorat, Univ. 7 novembre Carthage (Tunisie), 227 p., 129.
- Box, G. & Jenkins, G.M. (1976) *Time series analysis: Forecasting and control.* Holden Day Inc. Oakland, CA. 575 p.
- Buishand, T.A. (1982) Some methods for testing the homogeneity of rainfall records. *Journal of hydrology* 58:11-27.
- Buishand, T.A. (1984) tests for detecting a shift in the mean of hydrological.
- Canceill, M. (1974) *Régressions et corrélations en hydrologie : résultats élémentaires, règles pratiques.* Service Géologique National.
- Chambaz, H. (1990) *Contribution à l'étude de la méthode DPFT- différence première de la fonction de transfert- Application au bassin versant de l'oued Dis.* Thèse de magister, Ecole Nationale Polytechnique d'Alger (ENP), 200 p.
- Chatfield, C. (1989) *The Analysis of Time Series: An Introduction.* Chapman & Hall edition.
- Chiheb, F. (2011) *Quantification du transport solide et protection du bassin versant de Koudiat Acerdoune,* p49.
- Chu, T.W., Shirmohammadi, A., Montas, H. and Sadeghi, A. (2004) Evaluation of the SWAT Model Sediment and Nutrient Components in the Piedmont Physiographic Region of Maryland. *Transactions of the ASAE*, 47, 1523-1538. <https://doi.org/10.13031/2013.17632>
- Coulibaly, P., Anctil, F., Bobee, B. (1999) *Prévision hydrologique par réseaux de neurones artificiels : Etat de l'art.* *Journal of civil engineering*, 26; 293-301.
- Dagnélie, P. (1975) *Analyse statistique à plusieurs variables.* Les presses agronomiques de Gembloux.
- Dawes, W., Hatton, T. J. (Thomas J., and Resources, C. D. of W. (1993) *TOPOG IRM / by W. Dawes and T. Hatton.* CSIRO Division of Water Resources.
- Dechemi, N., Benkaci, T., Issolah A.(2003) *Modélisation des débits mensuels par les modèles conceptuels et les systèmes neuro-flous.* *Rev. Sci. Eau*, 16, 407-424.
- Dechemi, N. et Chambaz, H. (1994) *La prévision des crues du bassin versant de l'oued Dis (Sebbaou) par la méthode DPFT.* *Revue Sciences de l'eau*, 7; 331-342.
- Dimopoulos, I., Lek, S., Lauga, J. (1996) *Modélisation de la relation pluie-débit par les réseaux connexionnistes et le filtre de Kalman.* *Journal des sciences hydrologiques*, 41; 177-193.

Références bibliographiques

- Eckert, N. (2002) Prise en compte des couverts neigeux temporaires au sein d'un modèle pluie débit.
- French, M., Krajewski, W., Cuykendall, R. (1992) Rainfall forecasting in space and time using neural network. *Journal of hydrology*, 137; 1-31.
- Gaume, E. (2002) Eléments d'analyse sur les crues éclaircies. Thèse de doctorat. Ecole nationale du génie rural, des eaux et des forêts (Paris ; Nancy)
- Gineste, P. (1997) Télédétection spatiale et modélisation du fonctionnement hydrologique des petits bassins versants. I. Problématique de la modélisation. 367-378.
- Grisollet, H. , Guilmet, B. , Arléry, R. (1972) Climatologie. Méthodes et pratiques, Paris : Gauthier-Villars, (coll. Monographies de météorologie) p. 51.
- Gumbel, E. (1958) : méthodes graphiques pour l'analyse des débits de crue. Conférence du 24/04/1956 – Houille Blanche n°5 ; novembre 1956. *Statistical theory of floods and drought. Journal of the institution of water engineers*, vol n°3, 3 may 1958.
- Hernandez, M., Nearing, M. A., Al-Hamdan, O. Z., Pierson, F. B., Armendariz, G., Weltz, M. A., Spaeth, K. E., Williams, C. J., Nouwakpo, S. K., Goodrich, D. C., Unkrich, C. L., Nichols, M. H., and Holifield Collins, C. D. (2017). The Rangeland Hydrology and Erosion Model : A Dynamic Approach for Predicting Soil Loss on Rangelands: DYNAMIC RANGELAND SOIL EROSION MODEL. *Water Resources Research*, 53(11), 9368-9391. <https://doi.org/10.1002/2017WR020651>
- Hsu, K., Gupta, H., Sorooshian, S. (1995) Artificial neural network Modelling of the rainfall-runoff process. *Water resources research*, 31; 2517- 2530.
- Hubert, P. (1997) Change points in meteorological time series. Extrait de “applications of time series analysis in astronomy and meteorology. Chapman and Hall édition, Londres pp 399-412.
- Hubert, P., Carbonnel, J-P. et Chaouche, A. (1989) Segmentation des séries hydrométéorologiques – application à des séries de précipitation et de débits de l’Afrique de l’Ouest. *Journal of hydrology* 110, 349-367.
- Hubert, P., Servat, E., Paturel, J-E., Kouame, B., Bendjoudi, H., Carbonnel, J-P et Lubes, H. (1998) La procédure de segmentation, dix ans après. *Water Resources in Africa during the XXth century. IAHS publication N° 252*, pp 267-273.
- Hubert, P. & Carbonnel, J. P. (1993) Segmentation des séries annuelles de débits des grands fleuves africains. *Bull, de liaison du CIEH no. 92, Ouagadougou, Burkina Faso*, 3-10.
- Kirkby, M. J., Naden, P. S., Burt, T., and Butcher, D. P. (1993) *Computer Simulation in Physical Geography*, 2nd Edition (2nd edition). Wiley.
- Kingumbi, A. (2006) Modélisation hydrologique d'un bassin affecté par des changements d'occupation. Cas du Merguellil en Tunisie Centrale. Thèse de Doctorat, Université de Tunis El Manar.

Références bibliographiques

- Kotz, S., Johnson, N.L., Read, C.B. (1981) Encyclopedia of statistical sciences, New York, John Wiley, vol.1, 197-205, vol.8, 157-163, vol.9, 244-255.
- Larabi, S. (2017) Nouvelles approches de calibration automatique appliquées à un modèle hydrologique conceptuel semi-distribué. [Phd, Université du Québec, Institut national de la recherche scientifique]. <https://espace.inrs.ca/id/eprint/6650/>
- Lee, A.S.F., Heghinian, M., (1977) A Shift of the Mean Level in a Séquence of Independent Normal Random Variables: A Bayesian Approach. *Technometrics*, 19(4), 503-506.
- Lubes, H., Masson, J.M., Servat, E., Paturel, J.E., Boyer, J.F0 (1994) Caractérisation de fluctuations dans une série chronologique par application de tests statistiques. Etude bibliographique, Rapport 3, Programme ICCARE. ORSTOM, Montpellier.
- MacMillan, R., Furley, P., and Healey, R. (1993) Using hydrological models and geographic information systems to assist with the management of surface water in agricultural landscapes. *Landscape ecology and geographic information systems.*, 181-209.
- Maison, P. (2000) Un modèle hydrologique de suivi de la pollution diffuse en bassin versant (approche mécaniste simplifiée de la zone saturée), (Thèse INP Toulouse, France).
- Makhlouf, Z. (1994) Compléments sur le modèle pluie-débit GJ et essai d'estimation de ses paramètres. Thèse de Docteur es Sciences. Orsay: Université de Paris XI ORSAY, CEMAGREF d'Antony, 17 p.
- Mann, H.B., Withney, D.R. (1947) On a test of whether one of two random variables is stochastically larger than the other. *Annals of mathematical statistics* 18 : 50-60.
- Minns, A. and Hall, J. (1996) Artificial neural network as rainfall-runoff Models. *HydrologicalSciences journal*, 41; 399-417.
- Moriasi, D.N, M. W. Gitau, M.W., Pai, N., P. Daggupati, P. (2015) Hydrologic and water quality models : Performance measures and evaluation criteria; American society of Agricultural and biological Engineers, ISSN 2151-0032 , DOI 10.13031/trans 58.10715 ; p 1763-1785.
- Mouelhi, S., (2003) Vers une chaîne cohérente de modèles pluie-débit conceptuels globaux aux pas de temps pluriannuel, annuel, mensuel et journalier. Thèse de Doctorat, ENGREF, Cemagref Antony, France, 323 pp. 16.
- Mouelhi, S., Michel, C., Perrin, C. et Andreassian, V., (2006a) Linking stream flow to rainfall at the annual time step: the Manabe bucket model revisited. *Journal of Hydrology* 328, 283-296, doi:10.1016/j.jhydrol.2005.12.022.
- Mouelhi.S, MicheI C., Perrin C., Andreassian.V. (2006b) Stepwise development of a twoparameter monthly water balance Model, *J.Hydrol.*, Vol.318, Issues 1-4, 200-214.
- Morin, G., Paquet, T.P. (1995) Le modèle de simulation de quantité et de qualité CEQUEAU, guide de l'utilisateur, version 1.0B windows – Rapport de recherche n° 435, INRS-Eau, Québec.

Références bibliographiques

- Nash, J.E., Sutcliffe, J.V. (1970) River flow forecasting through conceptual models Part I—a discussion of principles. *J Hydrol* 27(3):282–290
- Neitsch, S. L. (2005) SWAT2005 theoretical documentation. <http://swatmodel.tamu.edu/media/1292/SWAT2005theory.pdf>.
<https://ci.nii.ac.jp/naid/10027756261/>
- Ogden, F. L., Garbrecht, J., DeBarry, P. A., and Johnson, L. E. (2001). GIS and distributed watershed models. II: Modules, interfaces, and models. *Journal of Hydrologic Engineering*, 6(6), 515-523.
- Özelkan, E. and Duckstein, L. (2001) Fuzzy conceptual rainfall–runoff Models. *Journal of hydrology*, 253, 41-68
- Quesney, A. (1999) Assimilation de mesures d’humidité de surface dans un modèle hydrologique conceptuel global. Apport de la télédétection radar ERS/SAR. Thèse de Doctorat, Université Paris 7, 350 p.
- Refsgaard, J. (1997) Parametrisation, calibration and validation of distributed hydrological models. *Journal of Hydrology* 198, p. 69-97.
- Renard, B. (2014) Probabilités et Statistiques appliquées à l’Hydrologie. Lecture-Type. France. pp.90. <https://hal.inrae.fr/hal-02597049>
- Rocha, N.R. (2014) Débits de crue et analyse hydrologique de petits bassins versants ; Mémoire de maîtrise ; Université Laval ; 115p
- Rodier, J.A . (1981) Phénomènes hydrologiques extrêmes - Sécheresses et crues exceptionnelles. Conférence internationale sur l’hydrologie et les bases scientifiques de la gestion rationnelle des ressources en eau. (Paris, 18-27 août 1981)
- Rodriguez, J.Y. (1989) Modélisation pluie-débit par la méthode DPFT. Thèse de Doctorat. INPG, Grenoble
- Serra Bertral, G. (1979) Cours d’Hydrologie de surface (1^{ère} partie). Institut National d’Agronomie, El Harrach, Algérie
- Shamseldin, Y. (1996) Application of neural network technique to rainfall-runoff Modeling. *Journal of hydrology*, 272-294.
- Smith, R., Goodrich, D., Woolhiser, D., and Unkrich, C. (1995). KINEROS-a kinematic runoff and erosion model. *Computer models of watershed hydrology.*, 697-732.
- Singh, V. P. (2018). Hydrologic modeling : Progress and future directions. *Geoscience letters*, 5(1), 1-18.
- Singh, V.P., and Woolhiser, D. (2002). Mathematical Modeling of Watershed Hydrology. *J. Hydrol. Eng.*, 7(4), 270-292. [https://doi.org/10.1061/\(ASCE\)1084-0699\(2002\)7:4\(270\)](https://doi.org/10.1061/(ASCE)1084-0699(2002)7:4(270))

Références bibliographiques

Singh, P., Ramasastri, K.S. ,. Singh, U.K. , Gergan, J.T. & DOBHAL, D.P. (1995) Hydrological characteristics of the Dokriani Glacier in the Garhwal Himalayas, Hydrological Sciences Journal, 40:2, 243-257, DOI: 10.1080/02626669509491407

Sneyers, S. (1975) : Note technique N° 143 sur l'analyse statistique des séries d'observations. OMM N° 415. Suisse. 189 pages.

Thompson, S. A. (2017). Hydrology for water management. CRC Press.

Weeks, D. & Boughton, W. C. (1987) Tests of ARMA Model forms for rainfall-runoff Modelling. Journal of hydrology, 91; 29-47.

Vazquez-Amábile, G.G. and Engel, B.A. (2005) Use of SWAT to Compute Groundwater Table Depth and Streamflow in the Muscatatuck River Watershed. Transactions of the ASAE, 48,991-1003. <http://dx.doi.org/10.13031/2013.18511>

WMO (1966) : Fifth Report of the WMO on the Advancement of Atmospheric Sciences and Their Application in the Light of Developments in Outer Space, WWW—Phase II.

Woolhiser, D. A. (1996). Search for physically based runoff model—A hydrologic El Dorado? Journal of Hydraulic Engineering. [https://doi.org/10.1061/\(ASCE\)0733-9429\(1996\)122:3\(122\)](https://doi.org/10.1061/(ASCE)0733-9429(1996)122:3(122))

Yair, A., and Lavee, H. (1985). Runoff generation in arid and semi-arid zones. Hydrological forecasting/edited by MG Anderson and TP Burt.

Zadeh L. (1965) : Fuzzy sets. Information and control, 8; 338-353.

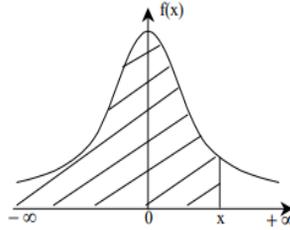
Zhu, T., and Xu, X. (2021). Simulation Modeling of Hydrological and Erosion Processes. In T. Zhu and X. Xu (Éds.), Watershed Erosion Processes (p. 147-155). Springer International Publishing. https://doi.org/10.1007/978-3-030-81151-8_9

Annexes

Annexe 01: Loi Normale centrée réduite

Probabilité de trouver une valeur inférieure à x .

$$F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-1/2u^2} du$$



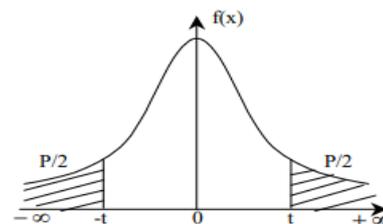
X	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767
2,0	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817
2,1	0,9821	0,9826	0,9830	0,9834	0,9838	0,9842	0,9846	0,9850	0,9854	0,9857
2,2	0,9861	0,9864	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,9890
2,3	0,9893	0,9896	0,9898	0,9901	0,9904	0,9906	0,9909	0,9911	0,9913	0,9916
2,4	0,9918	0,9920	0,9922	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936
2,5	0,9938	0,9940	0,9941	0,9943	0,9945	0,9946	0,9948	0,9949	0,9951	0,9952
2,6	0,9953	0,9955	0,9956	0,9957	0,9959	0,9960	0,9961	0,9962	0,9963	0,9964
2,7	0,9965	0,9966	0,9967	0,9968	0,9969	0,9970	0,9971	0,9972	0,9973	0,9974
2,8	0,9974	0,9975	0,9976	0,9977	0,9977	0,9978	0,9979	0,9979	0,9980	0,9981
2,9	0,9981	0,9982	0,9982	0,9983	0,9984	0,9984	0,9985	0,9985	0,9986	0,9986
3,0	0,9987	0,9987	0,9987	0,9988	0,9988	0,9989	0,9989	0,9989	0,9990	0,9990
3,1	0,9990	0,9991	0,9991	0,9991	0,9992	0,9992	0,9992	0,9992	0,9993	0,9993
3,2	0,9993	0,9993	0,9994	0,9994	0,9994	0,9994	0,9994	0,9995	0,9995	0,9995
3,3	0,9995	0,9995	0,9995	0,9996	0,9996	0,9996	0,9996	0,9996	0,9996	0,9997
3,4	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9998
3,5	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998

Table pour les grandes valeurs de x :

x	3	3,2	3,4	3,6	3,8	4	4,2	4,4	4,6	4,8
F(x)	0,99865003	0,99931280	0,99966302	0,99984085	0,99992763	0,99996831	0,99998665	0,99999458	0,99999789	0,99999921

Annexe 02: Loi de Student

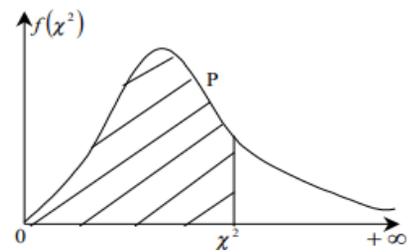
Valeurs de t ayant la probabilité P d'être dépassées en valeur absolue.



$n \setminus P$	90%	80%	70%	60%	50%	40%	30%	20%	10%	5%	1%
1	0,1584	0,3249	0,5095	0,7265	1,0000	1,3764	1,9626	3,0777	6,3137	12,7062	63,6559
2	0,1421	0,2887	0,4447	0,6172	0,8165	1,0607	1,3862	1,8856	2,9200	4,3027	9,9250
3	0,1366	0,2767	0,4242	0,5844	0,7649	0,9785	1,2498	1,6377	2,3534	3,1824	5,8408
4	0,1338	0,2707	0,4142	0,5686	0,7407	0,9410	1,1896	1,5332	2,1318	2,7765	4,6041
5	0,1322	0,2672	0,4082	0,5594	0,7267	0,9195	1,1558	1,4759	2,0150	2,5706	4,0321
6	0,1311	0,2648	0,4043	0,5534	0,7176	0,9057	1,1342	1,4398	1,9432	2,4469	3,7074
7	0,1303	0,2632	0,4015	0,5491	0,7111	0,8960	1,1192	1,4149	1,8946	2,3646	3,4995
8	0,1297	0,2619	0,3995	0,5459	0,7064	0,8889	1,1081	1,3968	1,8595	2,3060	3,3554
9	0,1293	0,2610	0,3979	0,5435	0,7027	0,8834	1,0997	1,3830	1,8331	2,2622	3,2498
10	0,1289	0,2602	0,3966	0,5415	0,6998	0,8791	1,0931	1,3722	1,8125	2,2281	3,1693
11	0,1286	0,2596	0,3956	0,5399	0,6974	0,8755	1,0877	1,3634	1,7959	2,2010	3,1058
12	0,1283	0,2590	0,3947	0,5386	0,6955	0,8726	1,0832	1,3562	1,7823	2,1788	3,0545
13	0,1281	0,2586	0,3940	0,5375	0,6938	0,8702	1,0795	1,3502	1,7709	2,1604	3,0123
14	0,1280	0,2582	0,3933	0,5366	0,6924	0,8681	1,0763	1,3450	1,7613	2,1448	2,9768
15	0,1278	0,2579	0,3928	0,5357	0,6912	0,8662	1,0735	1,3406	1,7531	2,1315	2,9467
16	0,1277	0,2576	0,3923	0,5350	0,6901	0,8647	1,0711	1,3368	1,7459	2,1199	2,9208
17	0,1276	0,2573	0,3919	0,5344	0,6892	0,8633	1,0690	1,3334	1,7396	2,1098	2,8982
18	0,1274	0,2571	0,3915	0,5338	0,6884	0,8620	1,0672	1,3304	1,7341	2,1009	2,8784
19	0,1274	0,2569	0,3912	0,5333	0,6876	0,8610	1,0655	1,3277	1,7291	2,0930	2,8609
20	0,1273	0,2567	0,3909	0,5329	0,6870	0,8600	1,0640	1,3253	1,7247	2,0860	2,8453
21	0,1272	0,2566	0,3906	0,5325	0,6864	0,8591	1,0627	1,3232	1,7207	2,0796	2,8314
22	0,1271	0,2564	0,3904	0,5321	0,6858	0,8583	1,0614	1,3212	1,7171	2,0739	2,8188
23	0,1271	0,2563	0,3902	0,5317	0,6853	0,8575	1,0603	1,3195	1,7139	2,0687	2,8073
24	0,1270	0,2562	0,3900	0,5314	0,6848	0,8569	1,0593	1,3178	1,7109	2,0639	2,7970
25	0,1269	0,2561	0,3898	0,5312	0,6844	0,8562	1,0584	1,3163	1,7081	2,0595	2,7874
26	0,1269	0,2560	0,3896	0,5309	0,6840	0,8557	1,0575	1,3150	1,7056	2,0555	2,7787
27	0,1268	0,2559	0,3894	0,5306	0,6837	0,8551	1,0567	1,3137	1,7033	2,0518	2,7707
28	0,1268	0,2558	0,3893	0,5304	0,6834	0,8546	1,0560	1,3125	1,7011	2,0484	2,7633
29	0,1268	0,2557	0,3892	0,5302	0,6830	0,8542	1,0553	1,3114	1,6991	2,0452	2,7564
30	0,1267	0,2556	0,3890	0,5300	0,6828	0,8538	1,0547	1,3104	1,6973	2,0423	2,7500
40	0,1265	0,2550	0,3881	0,5286	0,6807	0,8507	1,0500	1,3031	1,6839	2,0211	2,7045
50	0,1263	0,2547	0,3875	0,5278	0,6794	0,8489	1,0473	1,2987	1,6759	2,0086	2,6778
60	0,1262	0,2545	0,3872	0,5272	0,6786	0,8477	1,0455	1,2958	1,6706	2,0003	2,6603
80	0,1261	0,2542	0,3867	0,5265	0,6776	0,8461	1,0432	1,2922	1,6641	1,9901	2,6387
100	0,1260	0,2540	0,3864	0,5261	0,6770	0,8452	1,0418	1,2901	1,6602	1,9840	2,6259
120	0,1259	0,2539	0,3862	0,5258	0,6765	0,8446	1,0409	1,2886	1,6576	1,9799	2,6174
200	0,1258	0,2537	0,3859	0,5252	0,6757	0,8434	1,0391	1,2858	1,6525	1,9719	2,6006
∞	0,1257	0,2533	0,3853	0,5244	0,6745	0,8416	1,0364	1,2816	1,6449	1,9600	2,5758

Annexe 03: Loi du χ^2

Valeurs de χ^2 a y a n t la probabilité P d'être dépassée



ddl/P	0,5%	1,0%	2,5%	5,0%	10,0%	50,0%	90,0%	95,0%	97,5%	99,0%	99,5%
1	0,000	0,000	0,001	0,004	0,016	0,455	2,706	3,841	5,024	6,635	7,879
2	0,010	0,020	0,051	0,103	0,211	1,386	4,605	5,991	7,378	9,210	10,597
3	0,072	0,115	0,216	0,352	0,584	2,366	6,251	7,815	9,348	11,345	12,838
4	0,207	0,297	0,484	0,711	1,064	3,357	7,779	9,488	11,143	13,277	14,860
5	0,412	0,554	0,831	1,145	1,610	4,351	9,236	11,070	12,832	15,086	16,750
6	0,676	0,872	1,237	1,635	2,204	5,348	10,645	12,592	14,449	16,812	18,548
7	0,989	1,239	1,690	2,167	2,833	6,346	12,017	14,067	16,013	18,475	20,278
8	1,344	1,647	2,180	2,733	3,490	7,344	13,362	15,507	17,535	20,090	21,955
9	1,735	2,088	2,700	3,325	4,168	8,343	14,684	16,919	19,023	21,666	23,589
10	2,156	2,558	3,247	3,940	4,865	9,342	15,987	18,307	20,483	23,209	25,188
11	2,603	3,053	3,816	4,575	5,578	10,341	17,275	19,675	21,920	24,725	26,757
12	3,074	3,571	4,404	5,226	6,304	11,340	18,549	21,026	23,337	26,217	28,300
13	3,565	4,107	5,009	5,892	7,041	12,340	19,812	22,362	24,736	27,688	29,819
14	4,075	4,660	5,629	6,571	7,790	13,339	21,064	23,685	26,119	29,141	31,319
15	4,601	5,229	6,262	7,261	8,547	14,339	22,307	24,996	27,488	30,578	32,801
16	5,142	5,812	6,908	7,962	9,312	15,338	23,542	26,296	28,845	32,000	34,267
17	5,697	6,408	7,564	8,672	10,085	16,338	24,769	27,587	30,191	33,409	35,718
18	6,265	7,015	8,231	9,390	10,865	17,338	25,989	28,869	31,526	34,805	37,156
19	6,844	7,633	8,907	10,117	11,651	18,338	27,204	30,144	32,852	36,191	38,582
20	7,434	8,260	9,591	10,851	12,443	19,337	28,412	31,410	34,170	37,566	39,997
21	8,034	8,897	10,283	11,591	13,240	20,337	29,615	32,671	35,479	38,932	41,401
22	8,643	9,542	10,982	12,338	14,041	21,337	30,813	33,924	36,781	40,289	42,796
23	9,260	10,196	11,689	13,091	14,848	22,337	32,007	35,172	38,076	41,638	44,181
24	9,886	10,856	12,401	13,848	15,659	23,337	33,196	36,415	39,364	42,980	45,558
25	10,520	11,524	13,120	14,611	16,473	24,337	34,382	37,652	40,646	44,314	46,928
26	11,160	12,198	13,844	15,379	17,292	25,336	35,563	38,885	41,923	45,642	48,290
27	11,808	12,878	14,573	16,151	18,114	26,336	36,741	40,113	43,195	46,963	49,645
28	12,461	13,565	15,308	16,928	18,939	27,336	37,916	41,337	44,461	48,278	50,994
29	13,121	14,256	16,047	17,708	19,768	28,336	39,087	42,557	45,722	49,588	52,335
30	13,787	14,953	16,791	18,493	20,599	29,336	40,256	43,773	46,979	50,892	53,672
31	14,458	15,655	17,539	19,281	21,434	30,336	41,422	44,985	48,232	52,191	55,002
32	15,134	16,362	18,291	20,072	22,271	31,336	42,585	46,194	49,480	53,486	56,328
33	15,815	17,073	19,047	20,867	23,110	32,336	43,745	47,400	50,725	54,775	57,648
34	16,501	17,789	19,806	21,664	23,952	33,336	44,903	48,602	51,966	56,061	58,964
35	17,192	18,509	20,569	22,465	24,797	34,336	46,059	49,802	53,203	57,342	60,275