

الجمهورية الجزائرية الديمقراطية الشعبية
REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE

وزارة التعليم العالي والبحث العلمي
MINISTERE DE L'ENSEIGNEMENT SUPERIEURE ET DE LA RECHERCHE SCIENTIFIQUE

CENTRE UNIVERSITAIRE MAGHNA

INSTITUT DES SCIENCES ET DE LA TECHNOLOGIE



المركز الجامعي - مغنية -

معهد العلوم والتكنولوجيا

EXTRAIT DU PROCES VERBAL DE LA REUNION DU CONSEIL SCIENTIFIQUE
DE L'INSTITUT DES SCIENCES ET DE LA TECHNOLOGIE

03 Juillet 2022

L'an deux mille vingtdeux, le Trois Juillet à 12h00se sont réunis les membres du Conseil Scientifique de l'Institut des Sciences et de la Technologie dont les noms suivent:

- **M. HOUARI Ahmed** : *Président du Conseil Scientifique ST* ;
- **M. BOUDGHENE STAMBOULI Yassine** : *directeur de l'Institut ST* ;
- **M. GUIZ Amine** : *Chef du département d'hydraulique, Institut ST* ;
- **Mme AINSEBA Nabila** : *Directrice adjointe de la poste graduation et de la recherchescientifique et des relations extérieures de l'Institut ST* ;
- **M. HAMRA Amine** : *Directeur-Adjoint des Etudes de l'Institut ST* ;
- **MmeBEKRI Yamina** : *Responsable de domaine ST* ;
- **MmeGHERISSI Radia** : *Responsable de la filière d'hydraulique, Institut ST* ;
- **Mme SARI Aouatef** : *Représentante du corps Maître de Conférences au Conseil Scientifique de l'Institut ST* ;
- **MmeBENMANSOUR Khadidja** : *Représentante du corps Maître de Conférences au Conseil Scientifique de l'Institut ST* ;
- **M. KAZI-TANI Hychem** : *Représentant du corps Maître de Conférences au Conseil Scientifique de l'Institut* ;
- **M. ZEGNOUNI Aymen** : *Représentant du corps Maître-Assistant au Conseil Scientifique de l'Institut ST* ;
- **M. GAOUR Imad** : *Représentant du corps Maître-Assistant au Conseil Scientifique de l'Institut ST* ;

Etaient absentes :

- **MmeSETTOUTI Nadéra** : *Représentante du corps Maître de Conférences au Conseil Scientifique de l'Institut ST* ;
- **MmeNEHARI Meriem** : *Responsable de la bibliothèque ST* ;



الجمهورية الجزائرية الديمقراطية الشعبية
REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE

وزارة التعليم العالي والبحث العلمي
MINISTERE DE L'ENSEIGNEMENT SUPERIEUREET DE LA RECHERCHE SCIENTIFIQUE

CENTRE UNIVERSITAIRE MAGHNIA

INSTITUT DES SCIENCES ET DE LA TECHNOLOGIE



المركز الجامعي - مغنية -

معهد العلوم والتكنولوجيا

Ordre du jour:

L'ordre du jour préétabli est le suivant:

1. Bilan pédagogique de l'année universitaire 2021/2022
2. Besoin de recrutement d'enseignants (2022/2023)
3. Besoin d'équipement des laboratoires
4. Etat d'avancement dans la préparation des Manifestation scientifiques
5. Dépôt des photocopiés pédagogiques
6. Confirmation des rapports d'expertise des photocopiés pédagogiques
7. Divers

6. Confirmation des rapports d'expertise des photocopiés pédagogiques :

Un seul photocopiépédagogique déposéauparavant auprès du Conseil Scientifique pour expertise a reçu des **rapports favorables pour sa publication**. Après examen des rapports d'expertise par le Conseil Scientifique, ce dernier a **confirmé les avis favorables accordés par les experts pour ce photocopié**. La fiche signalétique du photocopié en question est la suivnate :

N°	Intitulé	Spécialité	Auteur (e)
1	ANALYSE ET MODELISATION HYDROLOGIQUE : THEORIE ET APPLICATION	Hydraulique Urbaine	Dr. BELARBI Halima

L'ensemble des enseignants (es) ont été remerciés (es) pour leur présence et leur contribution. La séance fut levée à 13h30

Signé :

Président du Conseil Scientifique IST

المعهد الجامعي مغنية
الجلس العلمي
للمعهد
المركز الجامعي مغنية
د. موارى أحمد
رئيس المجلس العلمي

الجمهورية الجزائرية الديمقراطية الشعبية

République Algérienne Démocratique et Populaire

وزارة التعليم العالي والبحث العلمي

Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

المركز الجامعي - مغنية-

Centre Universitaire - Maghnia-



Institut des Sciences et de la Technologie

Département d'Hydraulique

Polycopie de cours destiné aux étudiants de 1^{er} année Master Hydraulique (Semestre 1)

ANALYSE ET MODELISATION HYDROLOGIQUE
THEORIE ET APPLICATION

Présenté par :

Dr. BELARBI Halima

Enseignant Chercheur

Maître de Conférences "B"

Hydraulique

Année Universitaire : 2019-2020

Avant-propos

En hydrologie, les statistiques interviennent à toutes les étapes des études hydrologiques, depuis le recueil des données et leur critique, pendant l'élaboration des modèles et jusqu'à l'évaluation de l'impact d'aménagement.

Ce polycopié est un support pédagogique, il répond au canevas ministériel de la matière analyse et modélisation hydrologique enseignée en première année Master Hydraulique Urbaine. L'aspect méthodologique est présent au cours et travaux dirigés de cette matière. Le document contient les trois chapitres du programme, chaque chapitre débute par des définitions, et des formules statistiques accompagnées d'une série d'exercices résolus.

Le chapitre I traite les notions de statistiques et de probabilités, il présente aussi les principales méthodes d'ajustement des données et des tests statistiques pour évaluer la qualité de ces ajustements ;

Le chapitre II est départagé en deux parties. Une première partie est consacrée à l'étude de la notion de corrélation et de la régression. La deuxième partie de ce chapitre présente les notions d'analyse des séries chronologiques et de traitement de données ;

Le chapitre III est une présentation sommaire de la modélisation hydrologique.

Pour l'élaboration de ce polycopié, nous avons utilisé de nombreux documents.

TABLE DES MATIERES

Avant-propos

LISTE DES FIGURES

LISTE DES TABLEAUX

LISTE DES SYMBOLES

Chapitre I

NOTIONS DE PROBABILITE ET D'ANALYSE STATISTIQUES APPLIQUEE A L'HYDROLOGIE

I.1. Rappels	1
I.1.1. Définitions et généralités.....	2
I.1.2. Analyse d'un échantillon	4
I.2. Types de fonctions de distribution (ou de densité) en hydrologie	13
I.2.1. Fréquence expérimentale au non dépassement (FND).....	14
I.2.2. Fréquence au dépassement (FD)	14
I.2.3. Choix de la formule pour la fréquence expérimentale	15
I.3. Valeurs extrêmes d'une variable	15
I.4. Distributions statistiques utilisées en hydrologie	16
I.4.1. Critères de choix d'une loi	17
I.4.2. Intervalle de confiance	17
I.4.3. Distribution gaussien (loi Normale ou de Gauss)	18
I.4.4. Loi de Galton (ou log-normale ou de Gibrat-Gauss)	21
I.4.5. Loi doublement exponentiel ou loi de Gumbel (ou Loi des Valeurs Extrêmes de type I)	23
I.4.6. Loi de Fréchet	26
I.4.7. Loi de Pearson III.....	26
I.5. Test d'ajustement	27
I.5.1. Généralité.....	27
I.5.2. Statistique Khi deux χ^2	27
I.5.3. Test de Kolmogorov -Smirnov	28
I.6. Mise en application	29

Chapitre II

CORRELATIONS ET ANALYSE DES DONNEES

II.1. Définitions	54
II.1.1. Notions de régression et de corrélation	54
II.1.2. Méthodes relatives à la régression : Cas de la régression à deux variables	56
II.1.3. Méthodes relatives à la corrélation.....	59

II.1.4. Complément statistique des données hydrologiques	60
II.2. Séries, variables et processus	62
II.2.1. Définitions	62
II.2.2. Valeurs caractéristiques d'une série chronologique.....	64
II.2.3. Lissage des séries chronologiques	66
II.3. Test d'homogénéité	67
II.3.1. Approche générale	67
II.4. Mise en application	69

Chapitre III MODELISATION HYDROLOGIQUE

III.1. Modélisation hydrologique	67
III.2. Critères de qualité d'un modèle	69
III.3. Critères de performance	70
III.4. Objet et enjeux de la modélisation pluie-débit	71
III.5. Calage, validation	72
III.6. Classification des modèles pluie-débit	73
III.6.1. Déterministe ou stochastique	73
III.6.2. Global, distribué ou semi-distribué.....	74
III.6.3. Empirique, conceptuel ou théorique.....	75

REFERENCES BIBLIOGRAPHIQUES

ANNEXES

LISTE DES FIGURES

Chapitre I NOTIONS DE PROBABILITE ET D'ANALYSE STATISTIQUES APPLIQUEE A L'HYDROLOGIE

Fig. I.1. Branches de la statistique.	2
Fig. I.2. Positions respectives du mode, de la médiane et de la moyenne (Goldfarb and Pardoux, 2013).	7
Fig. I.3. Mesure de la dissymétrie.	10
Fig. I.4. Mesure d'aplatissement.	11
Fig. I.5. Histogramme et polygone des fréquences des débits maximums annuels en m^3/s	13
Fig. I.6. Courbes des fréquences cumulées des débits maximums annuels d'un Oued en m^3/s	13
Fig. I.7. Fonction de répartition et de densité de probabilité.	14
Fig. I.8. Loi de Gauss (Variable de moyenne nulle et d'écart-type 1).	19
Fig. I.9. Loi de Gauss. Courbe et les points d'inflexion.	19
Fig. I.10. Fréquences cumulées croissantes et décroissantes des températures moyennes annuelles ($^{\circ}C$) de la station de Béni-Bahdel.	30
Fig. I.11. Histogramme des fréquences relatives des précipitations annuelles d'une station pluviométrique (Bassin versant de l'Oued Tafna).	33
Fig. I.12. Courbes des fréquences cumulées croissantes et décroissantes des totaux pluviométriques annuels d'une station pluviométrique (Bassin versant de l'Oued Tafna).	33
Fig. I.13. Détermination de la médiane graphiquement.	34
Fig. I.14. Détermination de la médiane graphiquement.	34
Fig. I.15. Histogramme et polygone des fréquences relatives des totaux pluviométriques annuels d'une station pluviométrique (Bassin versant d'Oued Fodda).	37
Fig. I.16. Courbes des fréquences cumulées croissantes et décroissantes d'une station pluviométrique (Bassin versant de l'Oued Fodda).	38
Fig. I.17. Portion des précipitations. Au moins de 410 mm (ligne rouge). Moins de 465 mm (ligne noir), entre 350 mm et 650 mm (lignes verte et bleu).	38
Fig. I.18. Les différentes probabilités comprises entre 680 mm et 700 mm.	39
Fig. I.19. Ajustement d'une loi normale. Papier de probabilité de Gauss.	42
Fig. I.20. Ajustement d'une loi normale. Papier millimétré.	42
Fig. I.21. Courbe enveloppe de l'intervalle de confiance à un seuil de confiance de 95%.	45
Fig. I.22. Ajustement d'une loi log- normale. Papier semi-logarithmique.	47
Fig. I.23. Ajustement d'une loi log- normale. Papier de probabilité log normale.	47
Fig. I.24. P_{max} de 15 min. Ajustement d'une loi de Gumbel. Papier millimétré.	51
Fig. I.25. P_{max} de 15 min. Ajustement d'une loi de Gumbel. Papier de probabilité de Gumbel.	51
Fig. I.26. P_{max} de 60 min. Ajustement d'une loi de Gumbel. Papier millimétré.	52
Fig. I.27. P_{max} de 60 min. Ajustement d'une loi de Gumbel. Papier de probabilité de Gumbel.	52
Fig. I.28. P_{max} de 120 min. Ajustement d'une loi de Gumbel. Papier millimétré.	53
Fig. I.29. P_{max} de 120 min. Ajustement d'une loi de Gumbel. Papier de probabilité de Gumbel.	54
Fig. I.30. Courbes Intensités – Durées – Fréquences.	54

Chapitre II CORRELATIONS ET ANALYSE DES DONNEES

Fig. II.1. Variation annuelle des lames précipitées et écoulées. Bassin versant d'Oued Sebdu. Période d'étude (1980- 2001).	54
Fig. II.2. Courbes de régression.	55
Fig. II.3. Représentation graphiques. Pluies à la station B en fonction à la station A.	70
Fig. II.4. Valeurs théoriques données par la droite de régression.	73
Fig. II.5. Méthode de double cumul. A station de référence et B Station à tester.	80
Fig. II.6. Méthode de double cumul. A station de référence et C Station à tester.	81
Fig. II.7. Méthode de double cumul. A station de référence et C Station à tester corrigée.	83
Fig. II.8. Méthode des cumuls des résidus. A station de référence et P1 Station à tester.	85

Fig. II.9. <i>Méthode des cumuls des résidus. A station de référence et P2 Station à tester.</i>	87
---	----

Chapitre III

MODELISATION HYDROLOGIQUE

Fig. III.1. <i>Représentation schématique de la modélisation pluie-débit.</i>	67
Fig. III.2. <i>Schématisme du processus d'identification d'un modèle.</i>	68
Fig. III.3. <i>Variables d'un modèle hydrologique.</i>	68
Fig. III.4. <i>Procédure (standard) simplifiée pour la modélisation de la relation pluie-débit.</i>	72
Fig. III.5. <i>Représentations schématiques des modèles hydrologiques globales, semi-distribué et distribué.</i>	74

LISTE DES TABLEAUX

Chapitre I

NOTIONS DE PROBABILITE ET D'ANALYSE STATISTIQUES APPLIQUEE A L'HYDROLOGIE

Tableau I.1. Effectifs des températures moyennes annuelles (°C) de la station de Béni-Bahdel.	29
Tableau I.2. Fréquences absolues et relatives des températures moyennes annuelles (°C) de la station de Béni-Bahdel.	30
Tableau I.3. Fréquences absolues et relatives cumulées des températures moyennes annuelles (°C) de la station de Béni-Bahdel.	30
Tableau I.4. Pluies mesurées, leurs logarithmes népériens et leurs inverses.	31
Tableau I.5. Pluies mesurées, leurs logarithmes népériens et leurs inverses.	32
Tableau I.6. Calcul des fréquences absolues et relatives d'une station pluviométrique (Bassin versant de l'Oued Tafna).	32
Tableau I.7. Courbes des fréquences cumulées croissantes et décroissantes des totaux. Pluviométriques annuels d'une station pluviométrique (Bassin versant de l'Oued Tafna).	33
Tableau I.8. Détermination de la médiane graphiquement.	35
Tableau I.9. Calcul des fréquences absolues, relatives et cumulées croissantes d'une station pluviométrique (Bassin versant de l'Oued Fodda).	36
Tableau I.10. Calcul des fréquences cumulées croissantes et décroissantes d'une station pluviométrique (Bassin versant de l'Oued Fodda).	38
Tableau I.11. Pluies mesurées, leurs carrées et l'écart à la moyenne carrée.	40
Tableau I.12. Ajustement d'une loi normale à la série de pluies annuelles recueillies à une station pluviométrique (Oued FODDA).	41
Tableau I.13. Application du test du χ^2	43
Tableau I.14. Ajustement d'une loi log- normale à l'échantillon à la série de pluies annuelles recueillies à une station pluviométrique (Oued FODDA).	46
Tableau I.15. Application du test du χ^2	48
Tableau I.16. Pmax de 15 min.	51
Tableau I.17. Pmax de 60 min.	52
Tableau I.18. Pmax de 120 min.	53
Tableau I.19. Pluies maximales et leurs intensités.	54

Chapitre II

CORRELATIONS ET ANALYSE DES DONNEES

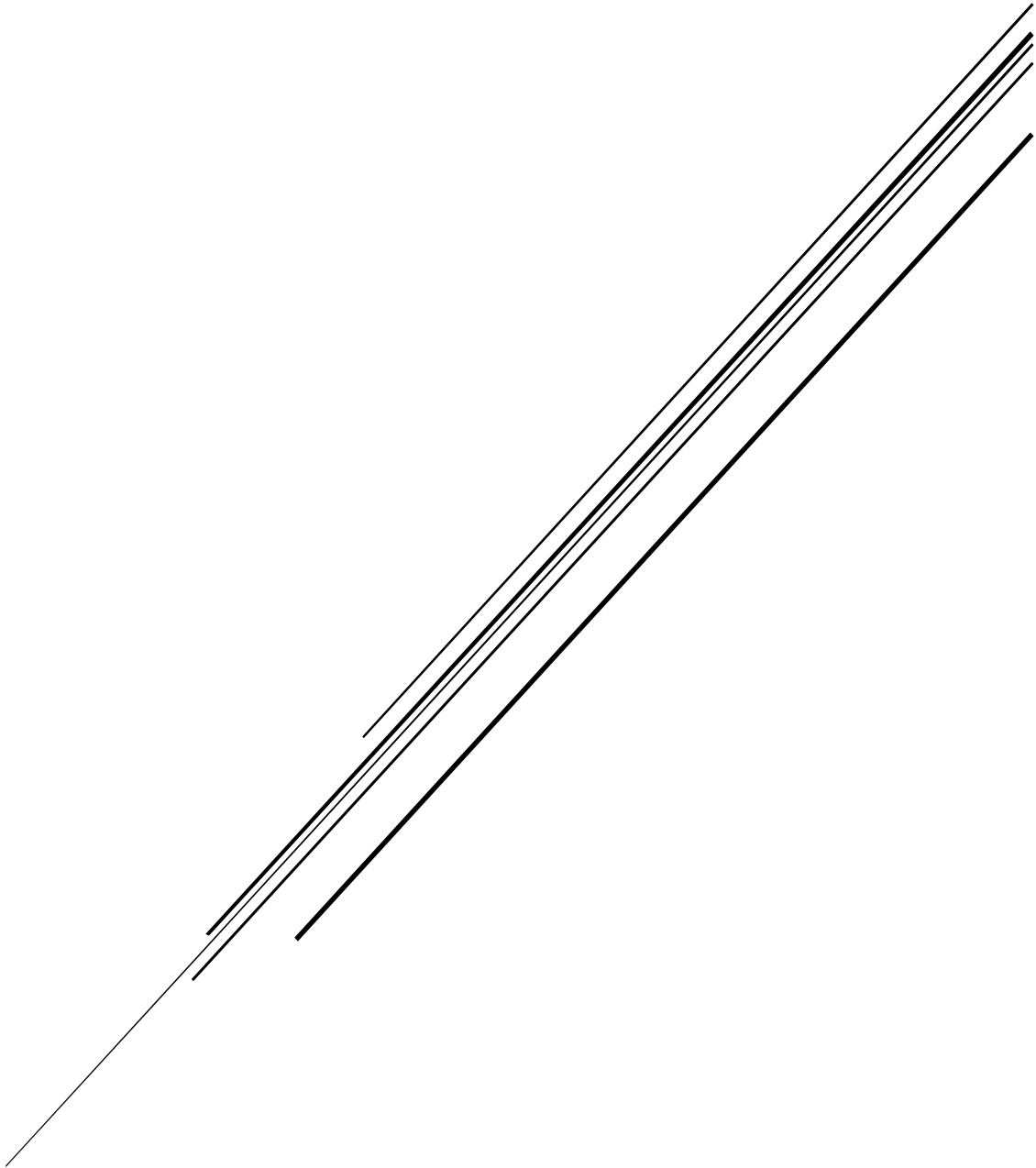
Tableau II.1. k couples de valeurs.	70
Tableau II.2. Calcul du coefficient de corrélation	70
Tableau II.3. Reconstituer les valeurs manquantes de la série B.	72
Tableau II.4. Valeurs théoriques de la droite de régression.	72
Tableau II.5. Test de Wilcoxon.	75
Tableau II.6. Test de Mann-Whitney.	77
Tableau II.7. Méthode de double cumul. A station de référence et B Station à tester	80
Tableau II.8. Méthode de double cumul. A station de référence et C Station à tester.	81
Tableau II.9. Méthode de double cumul. A station de référence et C Station à tester corrigée.	83
Tableau II.10. Méthode des cumuls des résidus. A station de référence et P1 Station à tester.	85
Tableau II.11. Méthode des cumuls des résidus. A station de référence et P2 Station à tester.	86

LISTE DES SYMBOLES

x_a, x_b	Bornes intervalle de variation de la variable aléatoire.
K	Coefficient d'aplatissement
C_s	Coefficient d'asymétrie
r	Coefficient de corrélation
R^2	Coefficient de détermination
CV	Coefficient de variance
β_1 et β_2	Coefficients de Pearson
R	Critère de Nash-Sutcliffe
Q_2	Deuxième quartile
$e_{\bar{x}}$	Ecart absolu moyen
S_x	Ecart type
E	Ensemble
\hat{r}	Estimation du coefficient de corrélation
ET	Etendue
EIQ	Etendue interquartile
e	Evénement
$f(u)$	Fonction de densité
$F(x)$	Fonction de répartition
f_{oi}	Fréquence absolue observée
FD	Fréquence cumulée au Dépassement
FND	Fréquence cumulée au Non Dépassement
f_i	Fréquence relative
f_{t1}	Fréquences théoriques
H_1	Hypothèse alternative
H_0	Hypothèse vraie ou nulle.
χ^2	Khi deux
Me	Médiane
Mo	Mode
m_k	Moments centrés
a_k	Moments non centrés
\bar{x}	Moyenne arithmétique
y_x	Moyenne conditionnelle
G	Moyenne géométrique
H	Moyenne harmonique
y_k	Moyennes mobiles impaire
T	Période de retour
R_t	Persistance
Q_1	Premier quartile
Pr	Probabilité
q	Probabilité de non apparition
α	Seuil de signification.
N	Taille de l'échantillon
TD_t	Tendance déterministe
TS_t	Tendance stochastique
Q_3	Troisième quartile
VA	Variable aléatoire
u	Variable réduite
X, Y	Variables statistiques
$V(X)$	Variance
C_t	Variation cyclique
Sa_t	Variation saisonnière
ε_t	Variations résiduelles

NOTIONS DE PROBABILITE ET D'ANALYSE STATISTIQUES APPLIQUEE A L'HYDROLOGIE

Chapitre I



NOTIONS DE PROBABILITE ET D'ANALYSE STATISTIQUES APPLIQUEE A L'HYDROLOGIE

Les phénomènes hydrologiques évoluent dans le *temps* et dans l'*espace* de manière partiellement **déterministe** (prévisible) et partiellement **aléatoire** (non prévisible). La composante aléatoire est parfois tellement prépondérante qu'il n'existe aucune corrélation entre les valeurs successives qui quantifient ces phénomènes. Ces observations sont donc indépendantes les unes des autres. De tels processus hydrologiques, sont dits **stochastiques**, et se prêtent bien à des analyses statistiques, particulièrement pour la description des **crues** et des **étiages**.

La description statistique des *crues* est particulièrement importante pour le dimensionnement des barrages, des ponts ou d'autres infrastructures hydrauliques. L'hydrologue doit choisir **une valeur de la crue de projet** et, par conséquent, il doit accepter *le risque* que cette valeur soit dépassée. *Le risque* apparaît donc comme une exposition à **un événement non désiré** qui pourrait détruire un ouvrage ou, tout au moins, l'endommager. Il s'exprime en **probabilité d'occurrence de l'événement** et dans la pratique, on peut en faire l'estimation par l'analyse statistique des fréquences des observations (*Anonyme, 2015a*).

Le but de ce premier chapitre est d'aborder cet **aspect statistique et probabilisé** de l'hydrologie. Après un bref aperçu des exigences imposées aux données hydrologiques. En prenant comme exemple les données hydrométriques, on présente des notions de **statistiques** et de **probabilités**. Suivent les fonctions de distribution de probabilité les plus couramment utilisées, ainsi que les principales méthodes d'ajustement des données. On propose ensuite deux tests statistiques pour évaluer la qualité de ces ajustements, puis on aborde le problème de l'incertitude provenant de ces distributions à l'aide des intervalles de confiance.

I.1. Rappels

Les statistiques sont une **science** qui utilise des méthodes scientifiques pour **collecter, organiser, synthétiser, présenter et analyser** les données. Elle permet aussi de tirer des **conclusions valables** et de **prendre des décisions** raisonnables sur la base de ces analyses.

Les statistiques permettent d'exploiter cette information pour établir toute relation de causalité par l'analyse et l'interprétation. On divise souvent l'étude de la statistique, au sens général, en deux parties (Fig. I.1) :

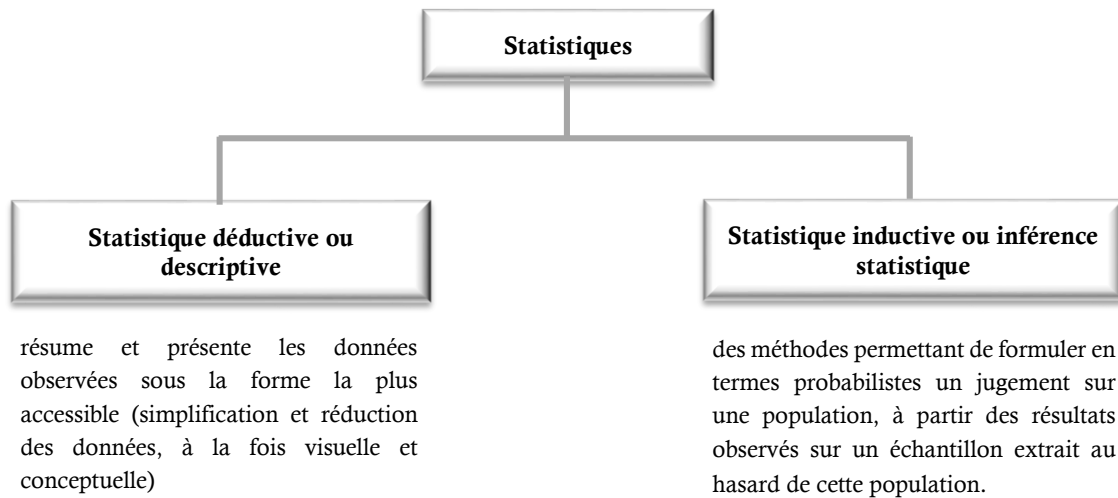


Fig. I.1. Branches de la statistique.

Il ne faut pas confondre *la* statistique qui est la science qui vient d'être définie et *une* statistique qui est un ensemble de données chiffrées sur un sujet précis (Vessereau, 1972).

I.1.1. Définitions et généralités

I.1.1.1. Répartition, échantillon et variable aléatoire

L'ensemble des événements fini ou infini est appelé *population*. Les éléments de la population sont appelés *individus* ou *unités statistiques* (ou *objet*).

Un *échantillon* est un sous ensemble de la population. La *taille* de l'échantillon est le nombre d'*événements* qui le constituent.

On dira qu'un échantillon est *exhaustif* lorsque sa taille est celle de la population. En hydrologie de surface, les populations sont généralement infinies et par conséquent, les échantillons sont *non exhaustifs* (Laborde, 2009).

La *série* des valeurs disponibles constitue un *échantillon* comprenant un nombre limité de valeurs appelées *observations*.

On appelle *variable aléatoire* (ou *variate*) notée par fois (V.A) le caractère numérique (ou non) qui distingue chaque événement de l'échantillon ou de la population. Notée en majuscule. Cette variable prendra une valeur à chaque tirage aléatoire, à chaque réalisation k (Bois et al., 2007).

Dans l'utilisation que nous ferons des statistiques, nous aurons toujours affaire à des *variables numériques discrètes* ou *continues*. Elles seront *discrètes* (on dit aussi *discontinue*) si les valeurs

numériques qu'elles peuvent prendre sont des valeurs isolées et bien déterminer (par exemple, nombre de jours de gel par an) (*Vessereau, 1972*). Par contre, on dira qu'une variable est *continue*, si ces valeurs numériques peuvent varier de façon continue (pluie journalière, débit, etc.). Ceci est le cas le plus fréquent, bien que l'on puisse remarquer que la précision des mesures limite la continuité de variation des valeurs numériques mesurées. Plus concrètement, si on mesure les pluies journalières au de millimètre, les pluies varieront de façon discrète avec un pas $\frac{1}{10}$ de millimètre. Par ailleurs, les valeurs numériques que peut prendre la variable X sont comprises entre deux bornes x_a et x_b finies ou infinies : $-\infty \leq x_a \leq x_b \leq +\infty$. L'intervalle est appelé *intervalle de variation de la variable aléatoire* (*Laborde, 2009*).

I.1.1.2. Notion de probabilité

Supposons que l'on puisse dénombrer l'ensemble E de toutes les éventualités également possibles et soit N le nombre de ces éventualités. Si parmi l'ensemble E, il n'existe que n éventualités favorables à un événement e, on dira que la probabilité de voir arriver l'événement e est la suivante (*Bois et al., 2007*) :

$$\Pr\left(\frac{e}{E}\right) = \frac{n}{N} \quad (\text{I-1})$$

Malheureusement pour nous, il est bien souvent impossible de dénombrer l'ensemble E et nous devons nous contenter d'une approche expérimentale. En effet, soit une série d'expériences à partir de laquelle on observe un intervalle E' (de taille N') d'éventualités. Parmi ces expériences, on observe n' réalisations de l'événement e, nous définirons alors la fréquence relative de l'événement e comme :

$$f = \frac{n'}{N'} \quad (\text{I-2})$$

En général, on peut admettre que si la taille N' augmente, la fréquence relative f se stabilise et tend sur la probabilité $\Pr\left(\frac{e}{E}\right)$:

$$\Pr\left(\frac{e}{E}\right) = \limite_{N' \rightarrow \infty} \frac{n'}{N'} \quad (\text{I-3})$$

I.1.1.3. Résultats fondamentaux

Dans ce qui suit, nous simplifierons l'écriture en remplaçant $\Pr\left(\frac{e}{E}\right)$ par $\Pr(e)$. La probabilité de réalisation de l'événement a parmi l'ensemble E, sachant que l'événement b est déjà réalisé, est notée : $\Pr\left(\frac{a}{b}\right)$.

▪ Si la réalisation de b exclut celle de a, on a : $\Pr\left(\frac{a}{b}\right) = 0$ (I-4)

▪ Si la réalisation de b entraîne nécessairement celle de a : $\Pr\left(\frac{a}{b}\right) = 1$ (I-5)

- Si on note (a + b) l'événement consistant en la réalisation de a ou de b, et (a · b) la réalisation de a et de b, nous aurons les relations suivantes :

$$\Pr(a + b) = \Pr(a) + \Pr(b) - \Pr(a \cdot b) \quad (\text{I-6})$$

$$\Pr(a \cdot b) = \Pr(a) \cdot \Pr\left(\frac{b}{a}\right) = \Pr(b) \cdot \Pr\left(\frac{a}{b}\right) \quad (\text{I-7})$$

I.1.2. Analyse d'un échantillon

En hydrologie, nous ne connaissons jamais la population totale mais nous ne disposons que d'échantillon de n valeurs d'une variable aléatoire X tirés dans cette population. Mais dès que n est grand (\geq quelques dizaines), la lecture du tableau de données n'est pas aisée, et il n'est pas utile de le transmettre entièrement pour permettre à un interlocuteur de s'en faire une idée. C'est pourquoi, il est intéressant d'effectuer une synthèse de ce tableau (*Bois et al., 2007*) :

- Synthèse numérique (on le résume en quelques chiffres) ;
- Synthèse graphique (on le résume en courbe) ;
- Synthèse analytique (on le résume par une fonction analytique, un modèle).

Certes on perdra de l'information mais on y gagnera en clarté.

I.1.2.1. Description numérique d'un échantillon

Une série de données peut être résumée par quelques valeurs numériques appelées *caractéristiques statistiques des séries*. La connaissance de celles-ci permet :

- De déduire certains aspects de la variable aléatoire considérée ;
- De choisir une loi de probabilité s'appliquant à cette série, dans le but de représenter le phénomène étudié.

Ces *caractéristiques statistiques des séries* sont classées en trois grandes catégories, à savoir :

- Caractéristiques de position (ou de tendance centrale) ;
- Caractéristiques de dispersion ;
- Caractéristiques de forme.

A. Paramètres de position (tendance centrale)

Elles donnent une idée de l'ordre de grandeur des valeurs constituant la série ainsi que la position où semblent se concentrer les valeurs de cette série. Les principales caractéristiques de tendance centrale sont (*Laborde, 2009 ; Bois et al., 2007 ; Goldfarb et Pardoux, 2013*) :

A.1. Moyennes

A.1.1. Moyenne Arithmétique

Soit la série statistique de données brutes : $x_1, \dots, x_j, \dots, x_n$, sa moyenne arithmétique \bar{x} a pour expression :

$$\bar{x} = \frac{(x_1 + x_2 + \dots + x_n)}{n} \Rightarrow \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (\text{I-8})$$

Propriétés

- La moyenne permet de résumer par un seul nombre la série statistique ;
- Elle prend en compte toutes les valeurs de la série et facile à calculer ;
- La somme des écarts à la moyenne est égale à zéro. $\sum_{i=1}^n (x_i - \bar{x}) = 0$

A.1.2. Moyenne géométrique

La moyenne géométrique G de la série de valeurs $x_1, \dots, x_j, \dots, x_n$ supposées toutes positives (strictement), est définie ainsi :

$$G = \sqrt[n]{\prod_{i=1}^n x_i} \Rightarrow \ln(G) = \frac{1}{n} \sum_{i=1}^n \ln(x_i) \quad (\text{I-9})$$

Lorsque la distribution de la variable statistique est donnée par les k couples (x_i, n_i) , les x_i étant tous positifs ; la moyenne géométrique G a pour expression :

$$G = \sqrt[n]{\prod_{i=1}^k x_i^{n_i}} = \prod_{i=1}^k x_i^{f_i} \Rightarrow \ln(G) = \sum_{i=1}^k f_i \ln(x_i) \quad (\text{I-10})$$

A.1.3. Moyenne harmonique

La moyenne harmonique H est l'inverse de la moyenne arithmétique des inverses des valeurs.

$$H = \frac{n}{\sum_{i=1}^n \frac{1}{x_i}} \text{ ou } H = \frac{1}{\sum_{i=1}^k \frac{f_i}{x_i}} \quad (\text{I-11})$$

A.2. Mode ou valeur dominante M_o

C'est la valeur M_o autour de laquelle on trouve *le plus de valeurs*, celle qui est *la plus fréquente*, ou *la plus probable*. On notera aussi que d'un point de vue analytique, le mode correspond au maximum de la densité de probabilité $f(x)$.

Propriétés

- Le mode a une signification concrète, qui rend son usage nécessaire dans un certain nombre de problèmes appliqués, par exemple la connaissance des heures de pointe dans la gestion des transports urbains ; les routes les plus fréquentées (points noirs dans le réseau routier), etc.
- Le mode n'existe pas toujours et quand il existe, il n'est pas toujours unique ;
- Si après regroupement des données en classes, on trouve deux ou plusieurs modes différents, on doit considérer que l'on est en présence de deux ou plusieurs populations distinctes ayant chacune leurs caractéristiques propres ; dans ce cas, la moyenne arithmétique n'est pas une caractéristique de tendance centrale.

A.3. Médiane M_e

La médiane M_e est plutôt une moyenne de position. Elle n'est qu'une forme particulière de quantile (appelés aussi **fractile**).

La médiane est la valeur, observée ou possible, dans la série des données classées par ordre croissant (ou décroissant) qui partage cette série en deux parties comprenant exactement le même nombre de données de part et d'autre de M_e .

Propriétés

- Le calcul de la médiane est rapide ;
- La médiane n'est pas influencée par les valeurs extrêmes ou aberrantes ;
- La médiane est influencée par le nombre des données mais non par leurs valeurs, elle ne peut donc pas être utilisée en théorie de l'estimation ;
- Si la variable statistique est discrète, la médiane peut ne pas exister ; elle correspond seulement à une valeur possible de cette variable ;
- La médiane est le point d'intersection des courbes cumulatives croissantes et décroissantes ;

Si la distribution des valeurs est symétrique, ces *trois caractéristiques de tendance* centrales sont *confondues* (Fig. I.2).

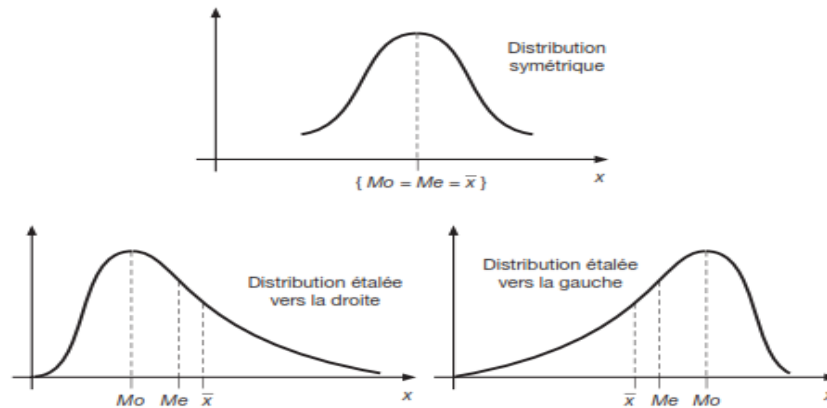


Fig. I.2. Positions respectives du mode, de la médiane et de la moyenne (Goldfarb et Pardoux, 2013).

En règle générale, lorsqu'il y a une dissymétrie marquée de la distribution statistique, la médiane est généralement préférable à la moyenne car elle est moins influencée par les valeurs exceptionnelles qui sont à l'origine de la dissymétrie. C'est seulement lorsque la dissymétrie est peu marquée, qu'on pourra utiliser la moyenne.

A.4. Quantiles

Le quantile d'ordre α ($0 \leq \alpha \leq 1$), noté x_α , est tel qu'une proportion α des individus ait une valeur du caractère X inférieure ou égale à x_α .

On utilise couramment les quantiles d'ordre 1/4, 1/2 et 3/4. Ils sont ainsi notés et nommés :

- $Q_1 = \text{premier quartile} = x_{0,25}$
- $Q_2 = \text{deuxième quartile} = \text{médiane} = x_{0,5}$
- $Q_3 = \text{troisième quartile} = x_{0,75}$

B. Paramètres de dispersion

Après avoir positionné la gamme de valeurs de X, on cherche à donner une idée de la fluctuation des x_i dans l'échantillon. Parmi les paramètres de dispersion les plus employés on cite :

B.1. Etendue (intervalle de variation)

L'étendu d'une série statistique est la différence entre la plus grande valeur et la plus petite valeur de la série.

$$ET = \text{étendue} = x_{\max} - x_{\min} \quad (\text{I-12})$$

Cet intervalle peut donner une idée fautive de la dispersion dans le cas où les valeurs extrêmes sont aberrantes, d'autre part, il n'est pas indépendant de l'effectif utilisé, des valeurs extrêmes risquant d'avantage d'apparaître au sien d'un grand effectif que parmi un contingent plus limité.

B.2. Etendue interquartile

L'étendue ou l'intervalle interquartile $[Q_1, Q_3]$ contient 50% des observations. Sa longueur, notée EIQ (Etendue Inter- Quartile) est un indicateur de dispersion :

$$EIQ = Q_3 - Q_1 \quad (I-13)$$

Le calcul de l'étendue interquartile à l'avantage par rapport à celui de l'étendue d'écart des valeurs extrêmes. Plus généralement, les longueurs des fourchettes définies par les déciles extrêmes, les centiles extrêmes constituent des indicateurs de dispersion contenant respectivement 80% et 98% des observations.

B.3. Ecart type

L'écart-type S_x d'une variable statistique X est la mesure de dispersion la plus couramment utilisée. Algébriquement, il se définit comme la racine carrée de la variance, et la variance est *la moyenne arithmétique des carrés des écarts à la moyenne arithmétique* :

$$V(X) = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \Rightarrow S_x = \sqrt{V(X)} \quad (I-14)$$

B.4. Ecart absolu moyen

L'écart absolu moyen à la moyenne $e_{\bar{x}}$ est la moyenne arithmétique des valeurs absolues des écarts à la moyenne arithmétique :

$$e_{\bar{x}} = \frac{1}{n} \sum_{i=1}^n |x_i - \bar{x}| \quad (I-15)$$

L'écart absolu moyen à la médiane e_{M_e} est la moyenne arithmétique des valeurs absolues des écarts à la médiane ; on démontre que c'est le plus petit écart absolu moyen :

$$e_{M_e} = \frac{1}{n} \sum_{i=1}^n |x_i - M_e| \quad (I-16)$$

B.5. Coefficient de variance

Le coefficient de variation est une mesure de dispersion relative, il permet d'apprécier la représentativité de la moyenne arithmétique par rapport à l'ensemble des observations (souvent exprimé en %) :

$$CV = \frac{S_x}{\bar{x}} \quad (I-17)$$

Il caractérise numériquement l'importance de la dispersion autour de la moyenne.

C. Paramètres de forme

Pour caractériser la forme d'une distribution, c'est-à-dire pour préciser l'allure de la courbe des fréquences relativement à son asymétrie et à son aplatissement, nous utiliserons des paramètres calculés à partir de valeurs appelées *moments*, dont la moyenne arithmétique et la variance, par exemple, sont des cas particuliers (*Touaibia, 2004*).

C.1. Moments

Les moments d'ordre K par rapport à une valeur C d'une série statistique sont définis comme :

$$a_k = \frac{1}{n} \sum_{i=1}^n (x_i - C)^k \quad (\text{I-18})$$

En pratique, on utilise :

Les moments non centrés ($C=0$) ou moment par rapport à l'origine:

$$a_k = \frac{1}{n} \sum_{i=1}^n (x_i)^k \quad (\text{I-19})$$

Les moments centrés ou moment par rapport à la moyenne ($C=\bar{x}$) :

$$m_k = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^k \quad (\text{I-20})$$

Remarque

Pour éviter les confusions, on notera par « a » les moments non centrés et par « m » les moments centrés ;

En pratique, le moment non centré d'ordre 1 se confond avec la moyenne $a_1 = \bar{x}$ pour $K=1$

Les moments centrés d'ordre 1 sont toujours nuls, $m_1 = 0$

Les moments centrés d'ordre 2 se confondent avec la variance, $m_2 = S_x^2$

D'une manière générale, les moments d'ordre pair m_4, m_6, \dots , sont comme la variance des paramètres de dispersion ;

Les moments centrés d'ordre impair m_3, m_5, \dots , sont des *indices de dissymétrie ou d'obliquité*. Ils sont nuls pour les distributions symétriques et différentes de zéro pour les distributions dissymétriques.

Les moments centrés d'ordre 3 et 4

Ils peuvent être déterminés d'une façon beaucoup plus simple :

$$m_3 = \frac{1}{n} \left[\sum_{i=1}^n x_i^3 - \frac{3}{n} (\sum_{i=1}^n x_i) (\sum_{i=1}^n x_i^2) + \frac{2}{n^2} (\sum_{i=1}^n x_i)^3 \right] \quad (I-21)$$

$$m_4 = \frac{1}{n} \left[\sum_{i=1}^n x_i^4 - \frac{4}{n} (\sum_{i=1}^n x_i) (\sum_{i=1}^n x_i^3) + \frac{6}{n^2} (\sum_{i=1}^n x_i)^3 (\sum_{i=1}^n x_i^2) - \frac{3}{n^3} (\sum_{i=1}^n x_i)^4 \right] \quad (I-22)$$

C.2. Coefficients de Pearson et de Fischer

Ils sont définis à partir des moments centrés d'ordre 2, 3 et 4

Les coefficients de Pearson β_1 et β_2 permettent de caractériser l'un le degré de dissymétrie et l'autre le degré d'aplatissement d'une distribution.



Karl Pearson
Mathématicien Britannique (1857 -1936)

Coefficients de Pearson

$$\beta_1 = \frac{m_3}{m_2^{3/2}} = \frac{m_3}{s^3} \quad (I-23) \quad \text{et} \quad \beta_2 = \frac{m_4}{m_2^2} = \frac{m_4}{s^4} \quad (I-24)$$

Mesure de la dissymétrie

Si $\beta_1 > 0$, la distribution est *étalée sur la droite*, on observe la succession mode, médiane, moyenne et **la dissymétrie est positive** (Fig. I.3) ;

Si $\beta_1 < 0$, la distribution est *étalée sur la gauche*, on observe la succession moyenne, médiane, mode et **la dissymétrie est négative** (Fig. I.3) ;

Si $\beta_1 = 0$, la distribution n'est pas forcément **symétrique** (mais si la distribution est symétrique, on a $\beta_1 = 0$) (Fig. I.3).

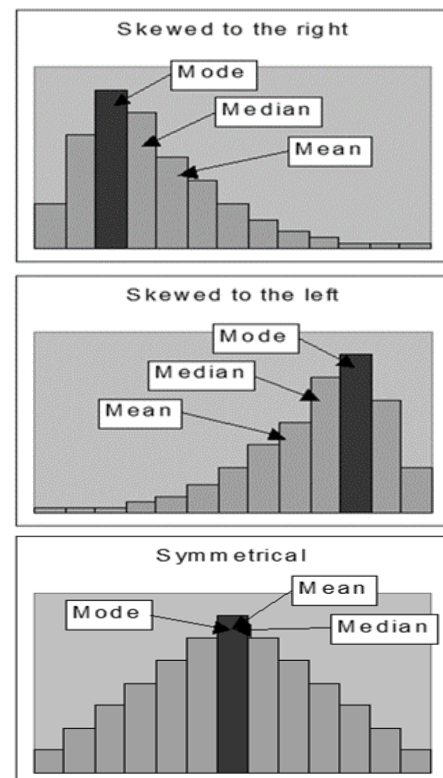


Fig. I.3. *Mesure de la dissymétrie (Anonyme, 2015a).*

Si $\beta_2 > 0$, la distribution est *moins aplatie* que la distribution normale, la distribution est dite **leptocurtique** (Fig. I.4) ;

Si $\beta_2 < 0$, la distribution est *plus aplatie* que la distribution normale, la distribution est dite **platicurtique** (Fig. I.4) ;

Si $\beta_2 = 0$, l'aplatissement est *même que pour la loi normale* et la courbe est dit **mésocurtique** (Fig. I.4).

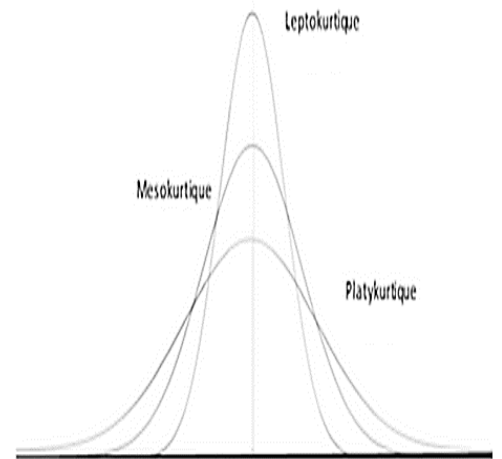


Fig. I.4. Mesure d'aplatissement (Anonyme, 2015a).

C.3 Coefficients de Pearson et de Fischer

Coefficients de Fischer

$$\gamma_1 = \frac{m_3}{m_2^{3/2}} = \frac{m_3}{\sigma^3} \quad (\text{I-25}) \quad \text{et} \quad \gamma_2 = \frac{m_4}{m_2^2} - 3 = \frac{m_4}{\sigma^4} - 3 \quad (\text{I-26})$$

Le coefficient d'asymétrie C_s n'est d'autre que γ_1 de Fischer, il peut être calculé d'une façon plus simple :

$$C_s = \frac{\sum_{i=1}^n (x_i - \bar{x})^3}{(n-1)\sigma^3} \quad \text{pour } n \leq 30 \quad (\text{I-27})$$

Le coefficient d'aplatissement appelé aussi *Kurtosis* K n'est autre que le coefficient β_2 de Pearson, il peut être calculé aussi par :

$$K = \frac{m_4}{m_2^2} = \frac{\sum_{i=1}^n (x_i - \bar{x})^4}{(n-1)\sigma^4} \quad (\text{I-28})$$



Ronald Fisher
Mathématicien Britannique (1890-1962)

I.1.2.2 Description graphique d'un échantillon

Une telle description se propose de déterminer quelle est la distribution statistique des observations de l'échantillon, c'est-à-dire (Anonyme, 2015a) :

- Quel est le pourcentage d'observations comprises entre deux valeurs (dans une classe limitée par ces deux valeurs) ;

Ou encore ;

- Quel est le pourcentage d'observations supérieures à une certaine valeur.

La représentation graphique d'une telle étude est un *histogramme des fréquences empiriques* qui permet d'en déduire un *histogramme des fréquences cumulées*.

A. Histogramme et polygone des fréquences empiriques

En général, les données sont brutes, elles ne sont pas organisées. Pour pouvoir tracer un histogramme on procède comme suit :

A.1 Ordonner la série

On peut ranger les valeurs étudiées soit dans l'ordre croissant ou dans l'ordre décroissant. Une valeur n'est inscrite qu'une seule fois et en face, dans un tableau, on indique le nombre de fois où l'on a observé cette valeur. Ce nombre est *l'effectif* de la valeur ou *fréquence absolue* (n_i). On peut également indiquer pour cette valeur *la fréquence relative* (f_i) qui est le *rapport entre la fréquence absolue de la valeur et le total des fréquences absolues (la taille de l'échantillon)*.

$$f_i = \frac{n_i}{N} \quad (\text{I-29})$$

A.2 Groupement des valeurs

Pour mettre en relief les caractéristiques de l'échantillon étudié, on opère des groupements en classes des valeurs. Dans la plupart des cas, on recherche un découpage en intervalles égaux. Pratiquement, pour trouver les distributions fréquentielles, l'on procède comme suit :

- On détermine la donnée la plus grande et la donnée la plus petite et on détermine l'étendue de l'échantillon.
- On divise cette étendue en un nombre convenable d'intervalles de même grandeur K' . On utilise souvent les formules suivantes (*Zendagui, 2011*):

$$K' = 1 + \frac{10 \times \log_{10} N}{3} \quad \text{Règle de Sturge} \quad (\text{I-30})$$

$$K' = 2,5 \sqrt[4]{N} \quad \text{Règle de Yule} \quad (\text{I-31})$$

Où, K' =nombre d'intervalles et N =La taille de l'échantillon

- On détermine le nombre d'observations (ou de données) dans chaque intervalle, c'est à dire, la fréquence absolue n_i de chaque classe.
- On détermine la fréquence relative $f_i = \frac{n_i}{N}$

L'ensemble des couples (x_i, n_i) ou (x_i, f_i) définit ce que l'on appelle *la fonction de distribution de la variable x*. Ainsi, *Un histogramme des fréquences* est une série de rectangles ayant (Fig. I.5) :

- leurs bases sur l'axe des x centrées au milieu des intervalles et dont la longueur est égale à la grandeur de l'intervalle ;
- leurs hauteurs sont égales aux fréquences.

La surface de chaque rectangle est proportionnelle à la fréquence de l'intervalle qu'il représente.

Le *polygone des fréquences* est obtenu en joignant les milieux des sommets des rectangles de l'histogramme. On complète la ligne polygonale aux moyens de segments, de façon telle que l'aire du polygone soit égale à l'aire de l'histogramme (Fig. I.5).

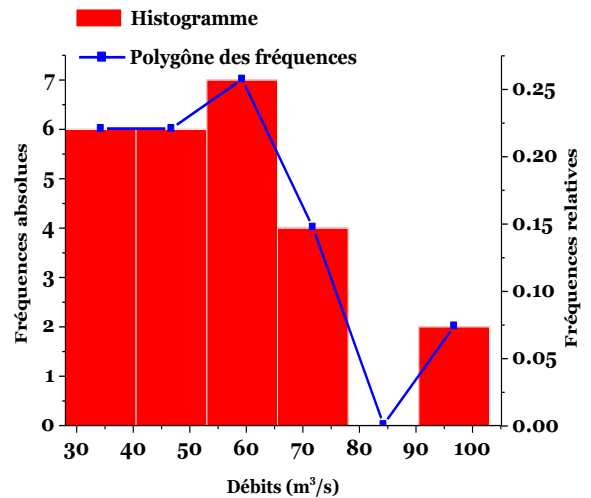


Fig. I.5. Histogramme et polygone des fréquences des débits maximums annuels en m³/s.

La représentation graphique de la distribution des fréquences relatives s'obtient à partir de l'histogramme ou du polygone des fréquences en changeant l'échelle verticale des fréquences absolues par les fréquences relatives (Fig. I.5).

B. Courbes des fréquences cumulées ou fonction de répartition

La *fonction de répartition*, dite aussi *fonction intégrale*, elle est constituée par l'ensemble des couples suivants :

x_i , cumul des fréquences de la plus petite valeur jusqu'à celle de x_i pour *les cumuls ascendants*, (de la plus petite à la plus grande valeur). C'est la *Fréquence cumulée au Non Dépassement (FND)* (Fig. I.6).

Ou ; x_i cumul des fréquences depuis celle de x_i jusqu'à celle de la plus grande valeur de l'échantillon pour *les cumuls descendants*. C'est la *Fréquence cumulée au Dépassement (FD)* (Fig. I.6).

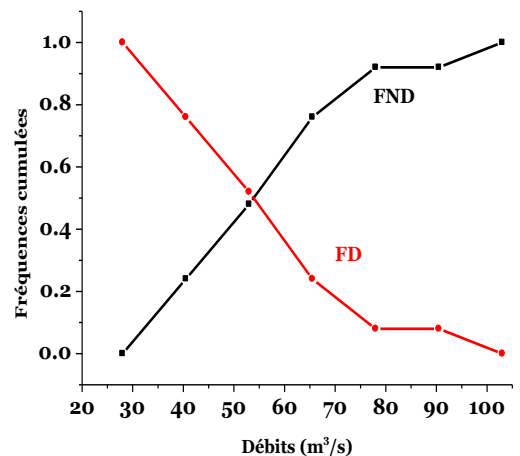


Fig. I.6. Courbes des fréquences cumulées des débits maximums annuels d'un Oued en m³/s.

I.2. Types de fonctions de distribution (ou de densité) en hydrologie

Un échantillon est considéré comme un sous-ensemble d'une population qui sera souvent infinie. Sur cette population, on peut définir une loi de probabilité : $F(x)$ où x correspond à une

valeur numérique. Cette *loi de probabilité*, ou *fonction de répartition* (Fig. I.7) qui exprime la *Probabilité que la Variable Aléatoire X reste inférieure ou égale à la valeur Numérique x*.

$$F(x) = \Pr(X \leq x) \quad (I-32)$$

La *fréquence f* d'un événement est le nombre de fois que cet événement se répète divisé par le nombre total d'événements observés. Lorsque le nombre total d'événements ou d'observations tend vers l'infini, cette mesure tend vers une valeur constante qui s'appelle la *probabilité (Pr)*.

En hydrologie, on parle indistinctement de *probabilité* et de *fréquence*.

Nous avons vu précédemment que l'ensemble des couples (x_i, n_i) ou (x_i, f_i) définit la *fonction de densité d'une variable* (Fig.I.7), c'est à dire : $f(x_i) = \frac{n_i}{N}$.

En pratique, on travaille soit à :

I.2.1. Fréquence expérimentale au non dépassement (FND)

Cette fréquence $F(x)$ correspond pour la population indéfinie à la probabilité au non dépassement. Si X est une variable aléatoire continue, susceptible de prendre une valeur quelconque sur la droite réelle, et x une de ces valeurs; la probabilité que X soit inférieure à x est (*Anonyme, 2015a*):

$$F(x) = \Pr(X < x) = \int_{-\infty}^x f(x)dx \quad (I-33)$$

Avec; $f(x)$: Densité de Probabilité et $F(x)$: Fonction de Répartition.

I.2.2. Fréquence au dépassement (FD)

Elle correspond à la probabilité au dépassement (*Anonyme, 2015a*)

$$F(x) = \Pr(X > x) = \int_x^{\infty} f(x)dx \quad (I-34)$$

- Fonction de répartition (probabilité au non dépassement théorique) $F(x)$ et $1 - F(x)$ (dépassement) ;

- Fonction de densité de probabilité $f(x) = \frac{d}{dx} F(x)$ (I-35)

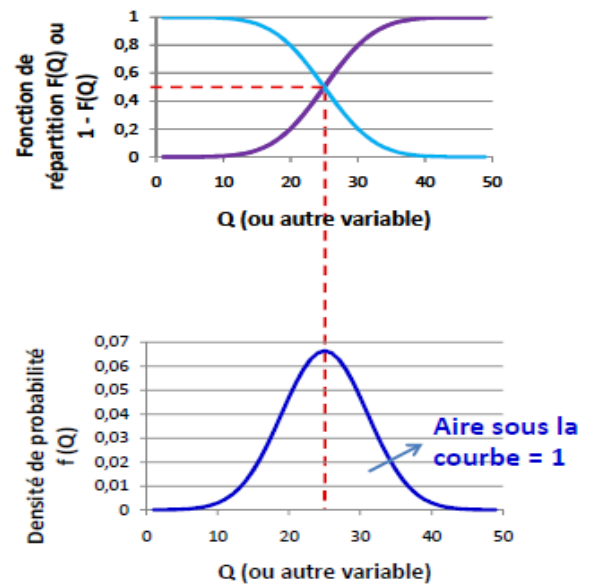


Fig. I.7. Fonction de répartition et de densité de probabilité (*Anonyme*).

$$\text{et, donc } F(x) = \int_{-\infty}^x f(x)dx \quad (\text{I-36})$$

En pratique, le traitement statistique d'une série hydrologique de N valeurs nécessite le calcul de la fréquence expérimentale. Son calcul nécessite un classement des valeurs par ordre croissant pour la fréquence au non dépassement et par ordre décroissement pour la fréquence au dépassement.

I.2.3. Choix de la formule pour la fréquence expérimentale

Il existe plusieurs formules de calcul pour la fréquence expérimentale, nous n'en citerons que quelques-unes des plus utilisées (*Touaibia, 2004*).

$$\text{Formule de Hazen pour les lois normales et quasi normales } F(x) = \frac{m-0,5}{N} \times 100 \quad (\text{I-37})$$

$$\text{Formule de Weibull } F(x) = \frac{m}{N+1} \quad (\text{I-38})$$

Formule de Tchegardayer pour les lois non normales

$$F(x) = \frac{m-0,3}{N+0,4} \quad \text{Si } C_s \geq 2CV \quad (\text{I-39})$$

$$F(x) = \frac{m-0,3}{N+0,2} \quad \text{Si } C_s < 2CV \quad (\text{I-40})$$

Avec; m: Numéro d'ordre; N: Taille de l'échantillon; C_s : Coefficient d'asymétrie et CV: Coefficient de variation.

I.3. Valeurs extrêmes d'une variable

Le calcul des valeurs extrêmes ou quantiles nécessite un ajustement à une loi de probabilité, après avoir calculé non seulement la fréquence expérimentale mais aussi les caractéristiques empiriques les plus importantes (moyenne arithmétique, écart type, coefficient de variation et d'asymétrie). Il faut choisir une loi de probabilité susceptible de s'ajuster d'une manière adéquate à la série hydrologique. Cette pratique revient donc à ajuster une loi théorique de répartition des fréquences à l'échantillon.

La probabilité d'apparition Pr ou f de l'évènement est liée à la *période de retour T*, dit aussi *intervalle moyen de récurrence*, défini par :

$$T = \frac{1}{Pr} \text{ ou } \frac{1}{f} \quad (\text{I-41})$$

Exemple

Un débit d'inondation dont la probabilité d'apparition ou de dépassement est 0,033 est appelé crue de 30 ans ($T = \frac{1}{0,033} = 30$ ans) car la probabilité est établie à l'aide des crues annuelles.

Les intervalles de récurrence recommandés pour le dimensionnement de certaines structures sont comme suit (*Anonyme*):

- Déversoirs de barrages où une crue dépassant la capacité peut mettre le barrage en danger et créer des dommages considérables et des pertes de vie : *500 à 1000 ans ou 10 000 ans.*
- Ponts sur routes importantes où l'exhaussement de l'eau crée par le pont peut entraîner des dommages importants ou la perte du pont : *50 à 100 ans.*
- Ponts sur routes secondaires ou ponceau sur routes importantes : *25 ans.*
- Ponceaux sur routes secondaires, égouts pluviaux, fossés de drainage : *5 à 10 ans.*
- Égouts pluviaux de moindre importance : *1 à 2 ans.*

Le calcul des probabilités permet de calculer quelles sont les chances de non apparition d'une valeur égale ou supérieure à X connaissant sa probabilité d'apparition au cours de l'année calculée par :

$$\text{Pr} = \frac{1}{T} \quad (\text{I-42})$$

La probabilité de non apparition se calcule par : $q = 1 - \text{Pr} = 1 - \frac{1}{T}$ (I-43)

I.4. Distributions statistiques utilisées en hydrologie

La probabilité est une application de l'espace des épreuves $[0 - 1]$. Le travail de l'hydrologue consiste à faire coïncider cette application avec le risque d'occurrence d'un événement bien défini. Les méthodes statistiques ou d'échantillonnage n'étudient qu'une grandeur d'un processus (pluie annuelle, le débit maximum, etc.) en excluant la notion de temps.

Les lois d'ajustement sont nombreuses et ne peuvent être appliquées à un échantillon que si les conditions *homogénéité* et *stationnarité* sont réunies parmi lesquelles (*Duband, 1982 ; Thiery, 1989; Touaibia, 2004; Touaibia et al. 2018*):

- Loi de Laplace-Gauss ou loi Normale
- Loi de Galton ou loi Log-Normale
- Loi de Fuller ou loi Exponentiel
- Loi Gamma, lois de Pearson I et II
- Loi de Gumbel ou loi double exponentielle
- Loi de Goodrich
- Loi de Fréchet

- Loi de Weibull....

Ces lois comportent 2 ou 3 paramètres estimés par différentes méthodes :

- Méthode des moments
- Méthode du maximum de vraisemblance
- Méthode du maximum d'entropie....

I.4.1. Critères de choix d'une loi

Ces critères de choix sont liés à un ajustement graphique d'abord et ensuite à un test de dispersion. L'allure des points sur du papier à probabilité permet à prime d'abord d'accepter ou de rejeter la loi (toute sinuosité, mauvaise courbure ou cassure de pente est considérée comme un mauvais ajustement).

Le test de dispersion permet d'accepter ou de rejeter l'appartenance d'une population à une loi donnée en fonction d'un seuil de signification fixé a priori. Cependant, une question délicate se pose : Quelle loi choisir lorsque plusieurs d'entre elles sont délicates ? Devant un tel problème, les critères de choix pouvant être retenues sont :

- A qualité égale la loi qui a le moins de paramètres est à retenir ;
- A qualité égale et nombre de paramètres égal, il faut comparer les estimations des deux lois. Si ces estimations sont voisines, il faut retenir la plus simple ; si elles sont nettement différentes, il faut prendre une nouvelle loi quitte à prendre un paramètre supplémentaire.

Remarque

Il est souhaitable de trouver une loi de 1 à 2 paramètres. Une loi de 3 paramètres ne peut être retenue que si sa qualité d'ajustement est nettement supérieure à celles des lois de 1 à 2 paramètres (ne jamais dépasser 3 paramètres sauf cas exceptionnel).

I.4.2. Intervalle de confiance

La fiabilité des estimations exprimée par l'intervalle de confiance sera en fonction de la taille de l'échantillon. L'intervalle de confiance est calculé aussi bien pour les caractéristiques de l'échantillon que sur les quantiles.

Bernier et Véron (1964) ont dressé, dans le cas de la loi de Gumbel, des tables permettant de calculer les intervalles de confiance à 70 et 95%.

Quelle que soit la loi théorique d'ajustement à un échantillon, le procédé de calcul est toujours le même. Il consiste à :

1. Classer les valeurs de l'échantillon par ordre croissant ou décroissant selon qu'on travaille à la fréquence au dépassement ou au non dépassement.
2. Affecter un numéro d'ordre aux valeurs classées ;
3. Calculer les caractéristiques empiriques de l'échantillon à savoir : moyenne, écart type ou coefficient de variation, coefficient d'asymétrie s'il existe...
4. Reporter les points sur papier à probabilité approprié à chaque loi ;
5. Tracer la droite ou la courbe théorique appropriée ;
6. Lire ou calculer la valeur extrême (quantile), c'est-à-dire la valeur correspondante à une probabilité donnée ;
7. Tester l'adéquation de la loi d'ajustement à l'aide d'un test numérique (test du Khi deux χ^2 , test de Kolmogorov-Smirnov,...) ;
8. Calculer l'intervalle de confiance, en général à 95%, sur les caractéristiques empiriques de l'échantillon et sur le ou les valeurs extrêmes.

I.4.3. Distribution gaussien (loi Normale ou de Gauss)

I.4.3.1. Généralité

Une variable appartient à la distribution normale lorsque sa fonction de répartition est de la

$$\text{forme : } F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^u e^{-\frac{u^2}{2}} du \quad (\text{FND}) \quad (\text{I-44})$$

u est appelée *variable réduite de Gauss*, qui a pour moyenne 0 et pour écart type 1. Elle est donnée

$$\text{par : } u = \frac{x - \bar{x}}{s_x} \quad (\text{I-45})$$

La transformation $x = \bar{x} + s_x u$ est l'équation de la droite de Henry sur papier à probabilité

$$\text{gaussienne. Sa densité de probabilité est } f(u) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{u^2}{2}} \quad (\text{I-46})$$

La dérivée première s'annule pour $u = 0$ et la dérivée secondaire pour $u = \pm 1$.

La fonction de densité est présentée par la courbe en *cloche symétrique* par rapport au mode $u = 0$ et avec des points d'inflexion pour $u = \pm 1$ (Fig. I.8).

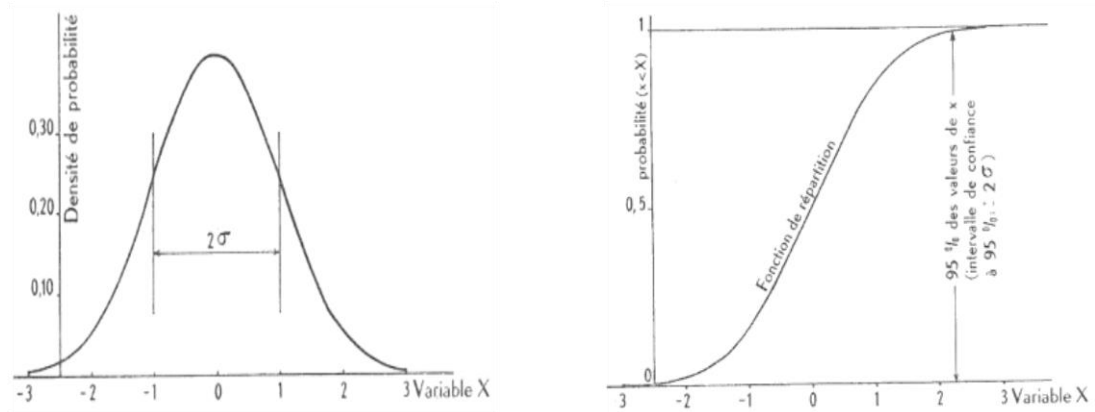


Fig. I.8. Loi de Gauss (Variable de moyenne nulle et d'écart-type 1) (Roche, 1963).

L'aire totale comprise entre la courbe et l'axe des abscisses mesure la probabilité pour que la variable u soit comprise entre $-\infty$ et $+\infty$.

Propriétés

- C'est une distribution symétrique par rapport à la moyenne ;
- La médiane est égale à la moyenne ;
- Elle est entièrement définie par deux paramètres : La *moyenne* et l'*écart type* ;
- L'aire totale sous la courbe est égale à un ;
- Les approches de la courbe normale ne touchent jamais l'axe des x et s'étendent plus loin de la moyenne ;
- Entre $\bar{x} - S_x$ et $\bar{x} + S_x$ (dans le centre de la courbe), le graphique s'incurve vers le bas. Il s'incurve vers le haut à gauche de $\bar{x} - S_x$ et à droite de $\bar{x} + S_x$. Les points où la courbe change vers le bas et vers le haut sont appelés des *points d'inflexion* (Fig. I.9).

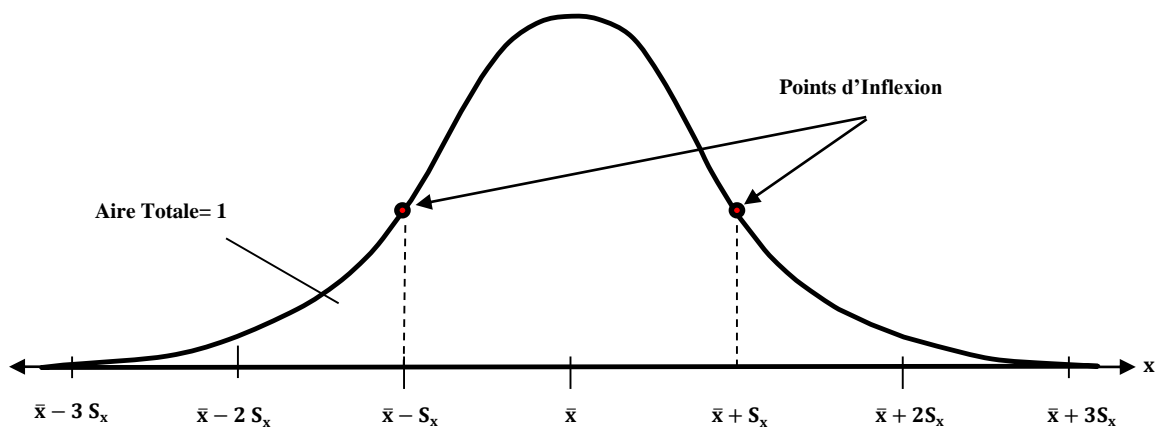


Fig. I.9. Loi de Gauss. Courbe et les points d'inflexion (Anonyme, 2015b).

I.4.3.2. Caractéristiques empiriques de l'échantillon

A. Méthode des moments

On obtient immédiatement les estimations suivantes :

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (\text{I-47})$$

$$\text{et } S_x = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2} \quad (\text{I-48})$$

B. Méthode du maximum de vraisemblance

Cette méthode conduit à une estimation de l'écart type

$$S_x = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - n\bar{x})^2} \quad (\text{I-49})$$

Elle est consistante mais moins correcte que la précédente que l'on préférera donc.

L'ajustement graphique consiste à reporter les valeurs en fonction de leurs fréquences expérimentales sur papier Gauss ou leurs variables réduites u sur papier millimétrée.

La droite de Henry nécessite la connaissance de deux points théoriques pour tracer. Connaissant \bar{x} et S_x , il suffit de prendre deux fréquences théoriques P_r de part et d'autre de la moyenne, de tirer u sur la table de Gauss (Annexe I) et de calculer la valeur correspondante par l'équation :

$$x_{P_r} = \bar{x} + S_x u_{P_r} \quad (\text{I-50})$$

L'estimation de l'intervalle de confiance pour une erreur α est calculée par la formule simplifiée :

$$\text{Pour la moyenne : } \bar{x} - u_{\frac{1-\alpha}{2}} \frac{S_x}{\sqrt{n}} < \mu < \bar{x} + u_{\frac{1-\alpha}{2}} \frac{S_x}{\sqrt{n}} \quad (\text{I-51})$$

$$\text{Pour l'écart type : } S_x - u_{\frac{1-\alpha}{2}} \frac{S_x}{\sqrt{2n}} < \sigma < S_x + u_{\frac{1-\alpha}{2}} \frac{S_x}{\sqrt{2n}} \quad (\text{I-52})$$

Pour un quantile x_p de probabilité p :

$$x_{P_r} - u_{\frac{1-\alpha}{2}} \frac{S_x}{\sqrt{2n}} \sqrt{u_{P_r}^2 + 2} < x_{P_r} < x_{P_r} + u_{\frac{1-\alpha}{2}} \frac{S_x}{\sqrt{2n}} \sqrt{u_{P_r}^2 + 2} \quad (\text{I-53})$$

On rappelle que μ et σ sont respectivement la moyenne et l'écart type de la population et que x_{P_r} est la valeur de la variable étudiée (pluie, débit, température, etc.) ayant une probabilité P_r de se réaliser.

I.4.4. Loi de Galton (ou log-normale ou de Gibrat-Gauss)

I.4.4.1. Généralité

Une variable aléatoire X a une distribution log normale lorsque $Y = \ln X$ est normale. La loi de Galton résulte de la loi normale mais rendue dissymétrie par un changement de variables. Sa

fonction de répartition est donnée par :
$$F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^u e^{-\frac{u^2}{2}} \quad (\text{FND}) \quad (\text{I-54})$$

1ère forme de la loi log normale

La variable réduite est de la forme :
$$u = \frac{\ln x - \overline{\ln x}}{S_{\ln x}} \quad (\text{I-55})$$

L'équation de la variable réduite présente sous la forme $\ln x = \overline{\ln x} + u S_{\ln x}$ (I-55), est l'équation d'une droite sur papier Gausso-logarithmique avec en abscisse l'échelle gaussienne et en ordonnée l'échelle logarithmique.

Procédé d'ajustement

- Classement des valeurs par ordre décroissant (fréquence au non dépassement) ou croissant (fréquence au dépassement) ;
- Calcul des caractéristiques empiriques de la série initiales \bar{x} et S_x ;
- Calcul des caractéristiques de la série transformée en logarithme $\overline{\ln x}$ et $S_{\ln x}$;
- Affecter un numéro d'ordre aux valeurs classées ;
- Calcul la fréquence expérimentale par une formule empirique ;
- Rapport des valeurs sur papier Gausso- logarithmique ou sur papier millimétré (une échelle arithmétique pour $\ln x = F(u)$);
- Détermination de la droite de Galton $\ln x = \overline{\ln x} + u S_{\ln x}$;
- Détermination de la valeur extrême soit graphiquement sur la droite soit analytiquement

$$x_{Pr} = e^{\ln x_{Pr}} = e^{\overline{\ln x} + u S_{\ln x}} \quad (\text{I-56})$$

2ème forme de la loi log normale

La variable réduite se trouve sous la forme :
$$u = \frac{\log(x-x_0)-a}{b} \quad (\text{I-57}),$$
 avec un intervalle de définition : $x \in]x_0, +\infty[$

La variable $\log(x - x_0)$ suit une loi de Gauss à 2 paramètres.

Avec ; a : moyenne des $\log(x - x_0)$ et b : écart type de $\log(x - x_0)$.

L'introduction du paramètre de position x_0 permet un ajustement de la loi log-normale lorsque la droite des fréquences expérimentales tracée sur le papier de Galton présente une concavité prononcée vers le haut on la corrige avec un x_0 négatif, si elle est vers le bas, elle est corrigée par un x_0 positif.

On passe de la loi de Gauss à la loi de Galton en faisant le changement de variable de x en $\log(x - x_0)$. Cette loi fait intervenir trois paramètres dont x_0 est le paramètre de position qui peut être déterminé graphiquement ou analytiquement, a et b sont des coefficients d'ajustement qui peuvent être déterminés par la méthode des moments ou la méthode de vraisemblance.

I.4.4.2. Estimations des paramètres

A. Méthode des moments

En utilisant la première formulation, les trois paramètres à déterminer sont x_0 , a et b . En posant :

Moyenne de x_i : \bar{x}

$$\text{Écart type des } x_i : S_x = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$$

$$\text{Moment centré d'ordre 3 : } m_3 = \frac{n}{(n-1)(n-2)} \sum_{i=1}^n (x_i - \bar{x})^3 \quad (\text{I-58})$$

On obtient trois équations à trois inconnues :

$$\left\{ \begin{array}{l} \frac{S_x^4}{m_3} = \frac{(\bar{x} - x_0)^3}{S_x^2 + 3(\bar{x} - x_0)^2} \end{array} \right. \quad (\text{I-59})$$

$$\left\{ \begin{array}{l} a = \frac{1,517}{\sqrt{\log \left[1 + \frac{S_x^2}{(\bar{x} - x_0)^2} \right]}} \end{array} \right. \quad (\text{I-60})$$

$$\left\{ \begin{array}{l} b = \frac{1,513}{a} - a \log(\bar{x} - x_0) \end{array} \right. \quad (\text{I-61})$$

L'équation (I-59) ne dépend que de x_0 et des valeurs prises dans l'échantillon. Elle ne peut être résolue que par approximations successives. Connaissant x_0 , on résout l'équation (I-60) qui donne a , puis connaissant a et x_0 on calcule b .

La méthode des moments présente le double inconvénient de nécessiter un calcul itératif et de donner parfois des valeurs de x_0 supérieures à la plus petite des valeurs rencontrées dans l'échantillon.

B. Méthode du maximum de vraisemblance

On obtient trois équations à trois inconnues x_0 , a et b :

$$\left\{ \begin{array}{l} 2,3026 \left[\sum_{i=1}^n \frac{1}{x_i - x_0} \right] \left[\frac{1}{n} \sum_{i=1}^n \log^2(x_i - x_0) - \frac{1}{n^2} \left(\sum_{i=1}^n \log(x_i - x_0) \right)^2 \right] = \frac{1}{n} \left[\sum_{i=1}^n \log(x_i - x_0) \right] \left[\sum_{i=1}^n \frac{1}{x_i - x_0} \right] - \sum_{i=1}^n \frac{\log(x_i - x_0)}{x_i - x_0} \end{array} \right. \quad (I-62)$$

$$a^2 = \frac{1}{\frac{1}{n} \sum_{i=1}^n \log^2(x_i - x_0) - \frac{1}{n^2} [\sum_{i=1}^n \log(x_i - x_0)]^2} \quad (I-63)$$

$$b = - \frac{a \sum_{i=1}^n \log(x_i - x_0)}{n} \quad (I-64)$$

Comme dans la méthode des moments, l'équation (I-62) ne dépend que de x_0 mais on doit la résoudre par approximations successives. Connaissant x_0 la formule (I-63) permet d'évaluer a, enfin, avec x_0 et a on détermine b par la formule (I-64). Cette méthode est préférable lorsque l'on dispose de moyens de calcul.

On passe de la loi de Gauss à la loi de Galton par le changement de variable x en $\log(x - x_0)$. Par conséquent, si on connaît x_0 , on pourra tracer l'équivalent de la droite d'Henry en portant $\log(x - x_0)$ en fonction de $F(x) = \frac{i-0,5}{n}$ sur un papier à probabilité gradué suivant les valeurs de la variable réduite de Gauss.

I.4.5. Loi doublement exponentiel ou loi de Gumbel (ou Loi des Valeurs Extrêmes de type I)

I.4.5.1. Généralités

C'est une distribution souvent utilisée pour représenter des valeurs extrêmes, c'est pourquoi on l'appelle parfois loi des valeurs extrêmes. Ce n'est cependant absolument pas un postulat et les valeurs extrêmes ne suivent pas forcément une distribution de Gumbel. Sa *fonction de répartition*, sous sa forme simple et généralement utilisée : $F(x) = e^{-\frac{x-x_0}{\alpha}}$ (I-65)

$F(x)$: Fréquence au dépassement de la valeur de x ;

x_0, α : Coefficients d'ajustement ;

Avec :

x_0 est le paramètre de position (mode) ;

α est le paramètre d'échelle différent de zéro et positif appelé aussi *gradex*.

Par un changement de variable $y = \frac{x-x_0}{\alpha}$, la loi de Gumbel s'écrit : $F(x) = e^{-y}$ (I-66)

$y = \frac{x-x_0}{\alpha}$ est la variable réduite de Gumbel

L'intervalle de variation de x est $x \in]-\infty, +\infty[$.

L'équation $y = \frac{x-x_0}{\alpha}$ présentée sous la forme $x = \frac{1}{\alpha} y + x_0$ est l'équation d'une droite qui représente la loi de Gumbel sur papier à probabilité Gumbel.

Ce papier porte en abscisse la fréquence au non dépassement et une échelle arithmétique de la variable réduite y .

L'ordonnée représente une échelle arithmétique de la variable étudiée.

1.4.5.2. Estimations des paramètres

A. Méthode des moments

La méthode des moments conduit à des estimations de calcul particulièrement simple et d'emploi très fréquent avec :

$$\alpha = 0,7797 S_x \tag{I-67}$$

$$x_0 = \bar{x} - 0,577 \alpha = \bar{x} - 0,444 S_x \tag{I-68}$$

$$\left(\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad \text{et} \quad S_x = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2} \right)$$

B. Méthode du maximum de vraisemblance

On aboutit au système résolutif suivant :

$$\left\{ \begin{array}{l} \bar{x} = \alpha + \frac{\sum_{i=1}^n x_i e^{-\frac{x_i}{\alpha}}}{\sum_{i=1}^n e^{-\frac{x_i}{\alpha}}} \end{array} \right. \tag{I-69}$$

$$\left\{ \begin{array}{l} x_0 = \alpha \ln \left[\frac{\sum_{i=1}^n e^{-\frac{x_i}{\alpha}}}{n} \right] \end{array} \right. \tag{I-70}$$

L'équation (I-69) ne peut se résoudre que par itération, ayant déterminé α , l'équation (I-70) permet la détermination de x_0 .

La méthode du maximum de vraisemblance ne s'utilise que lorsque l'on dispose de moyens de calcul automatique

C. Procédé d'ajustement

1. Classement des valeurs par ordre croissant en leur affectant un numéro d'ordre ;
2. Calculer la fréquence expérimentale ;
3. Calculer les caractéristiques empiriques de la série ;
4. Calculer la variable de Gumbel pour chaque valeur observée $y = -\ln(-\ln(F(x)))$; $\tag{I-71}$

5. Reporter les valeurs observées sur papier Gumbel ;
6. Calculer le coefficient de corrélation entre les valeurs observées et la variable de Gumbel dont la formule générale est

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} \quad (\text{I-72})$$

\bar{x} et \bar{y} sont respectivement les moyennes arithmétiques des variables X et Y.

- Si la corrélation est bonne, calculer les paramètres d'ajustement de la droite de Gumbel. La droite de régression ou droite de Gumbel est : $x = \frac{1}{\alpha}y + x_0$ où ; $\frac{1}{\alpha}$ est la pente de la droite et x_0 est l'ordonnée à l'origine.

y est la variable de Gumbel pour une probabilité donnée. Les paramètres $\frac{1}{\alpha}$ et x_0 peuvent être aussi déterminés par la méthode des moindres carrés ;

7. Tracer la droite de régression sur un papier Gumbel. Pour cela, il suffit de prendre deux valeurs de probabilité pour lesquelles, il faut déterminer la variable y qu'il suffira de remplacer dans la droite de régression.

Exemple ; $F(x) = 0,50$ on a $y = -\ln(-\ln(0,50)) = +0,36$

$F(x) = 0,80$ on a $y = -\ln(-\ln(0,80)) = +1,49$

En remplaçant, ces valeurs dans l'équation, nous retrouvons respectivement les valeurs probabilité 50% et 80%.

La valeur extrême ou quantile peut être déterminée graphiquement sur papier Gumbel ou analytiquement en utilisant simplement la droite de régression ou droite de Gumbel.

8. Calculer les intervalles de confiance en utilisant la formule :

$$\frac{\frac{y_\alpha}{\sqrt{n}} \sqrt{1 + 1,13U_{Pr} + 1,1u_{Pr}^2 \pm \frac{y_\alpha^2}{n}(1,1U_{Pr} + 0,57)}}{1 - 1,1\frac{y_\alpha^2}{n}} \quad (\text{I-73})$$

Avec :

- y_α est la variable réduite de Gauss correspondant à la fréquence au non-dépassement $1 - \frac{1-\alpha}{2}$;
- U_{Pr} est la variable réduite de Gauss correspondant à la fréquence au non dépassement Pr, ramenée à sa moyenne et à son écart type :

$$U_{Pr} = \frac{(-\ln(-\ln(F(x)))) - 0,577}{1,28} \quad (\text{I-74})$$

I.4.6. Loi de Fréchet

Parmi les distributions exponentielles généralisées, nous ne citerons que la loi de Fréchet qui est la plus fréquemment utilisée. Sous sa forme simplifiée, la fonction de distribution s'écrit :

$$F(x) = e^{-e^{-y}} \quad (\text{I-75})$$

Définie sur l'intervalle : $x \in]x_0, +\infty[$;

$$\text{Avec ; } y = \frac{\log(x-x_0)-\alpha}{\delta} \quad (\text{I-76})$$

Cette loi à trois paramètres :

x_0 : paramètre de position (mode) ;

δ : paramètre d'échelle (appelé aussi *gradex*) ;

α : paramètre de forme.

Elle dérive donc directement de la loi de Gumbel par le changement de variable de x en $\log(x - x_0)$. La fonction de distribution est alors à dissymétrie positive.

Dans la formulation citée plus haut, nous avons trois paramètres à déterminer. Dans ce cas, on est directement ramené à la loi de Gumbel par la transformation x en $\log(x - x_0)$ et par conséquent, on estimera : $\delta = 0,78 \sigma_{\log(x-x_0)}$ et $\alpha = \log(x - x_0) - 0,577\delta$ (I-77)

Les intervalles de confiance sont les mêmes que pour la loi de Gumbel mais au lieu de travailler sur x , on travaille sur $\log(x - x_0)$.

I.4.7. Loi de Pearson III

La forme standardisée de la loi de Pearson III est une transformation linéaire de la fonction gamma :

$$F(t) = K \int_t^\infty e^{\lambda t} (x + \sqrt{\lambda})^{\lambda-1} dt \quad (\text{I-78})$$

Avec :

$$K = \frac{e^{-\lambda\sqrt{\lambda}}}{\gamma(\lambda)} \quad (\text{I-79})$$

$\gamma(\lambda)$, fonction gamma et λ est un paramètre de forme ;

t est la variable réduite ayant pour expression :

$$t = \frac{x_i - \bar{x}}{S_x} = \frac{K_i - 1}{CV} ; \text{ avec : } K_i = \frac{x_i}{\bar{x}} \quad (\text{I-80})$$

S_x : Ecart type de la série ;

\bar{x} : Moyenne arithmétique de la série ;

K_i : Hydraulicité ou coefficient modulaire de Kotchérine.

Cette variable t n'est en fait que la fonction Rybkine Foster F_{Pr} qui est tabulée (Annexe II) et dépend du coefficient de variation CV , du coefficient d'asymétrie C_s et de la probabilité Pr .

$$t = F_{Pr} = \frac{K_{Pr}-1}{CV} \quad (I-81)$$

$$D'où: K_{Pr} = F_{Pr} \cdot CV + 1 \quad (I-82)$$

I.5. Test d'ajustement

I.5.1. Généralité

Lorsqu'on a procédé à un ajustement d'une loi de probabilité théorique, le problème qui se pose est de savoir si cette loi s'adapte. L'examen graphique ne peut suffire, il faut pouvoir tester par le calcul la qualité de l'ajustement réalisé. Ce test d'adéquation consiste à prendre une règle de décision concernant la validité d'une hypothèse relative à l'accord global d'une distribution empirique avec une distribution théorique spécifiée a priori ou ajustée sur les observations.

En général, l'application des tests d'adéquations consiste à vérifier l'hypothèse H_0 selon laquelle une certaine loi de répartition $F(x)$ s'ajuste à notre échantillon. C'est-à-dire que H_0 est considérée *hypothèse vraie* ou *nulle*. L'autre hypothèse est appelée *hypothèse alternative* H_1 .

Le risque consenti et choisi à l'avance et que nous appelons α de rejeter H_0 alors qu'elle est vraie est appelé *seuil de signification*. On a :

- α : probabilité de rejeter H_0 alors qu'elle est vraie, ou bien ;
- α : probabilité de nous tromper dans notre choix.

Dans ce cours, nous décrirons deux tests d'adéquations : le test du Khi deux χ^2 et le test de Kolmogorov-Smirnov.

I.5.2. Statistique Khi deux χ^2

La procédure d'utilisation de ce test est la suivante :

1. On divise l'intervalle de variation en k classes de façon que chacune d'elles contienne au minimum 5 données expérimentales (généralement, on s'abstient d'analyser des échantillons de moins de 20 valeurs). La classe i est borné par les valeurs x_{i-1} et x_i choisis arbitrairement. On détermine la fréquence absolue observée ou l'effectif de chaque classe : $f_{01}, f_{02}, \dots, f_{0k}$ avec $\sum f_{0i} = N$ et N est la taille de l'échantillon ;

2. On peut obtenir les fréquences théoriques : $f_{t1}=N p_1$; $f_{t2}=N p_2$, ..., $f_{tk}=N p_k$

f_{0k} avec p_i représente la probabilité que la variable étudiée prenne une valeur appartenant à la classe i ;

3. Pour évaluer l'ampleur de l'écart entre les fréquences absolues observées f_{0i} est les fréquences théoriques f_{ti} obtenues à partir de la loi théorique que l'on suppose adéquate, on utilise la

$$\text{quantité : } \chi^2 = \frac{(f_{01}-f_{t1})^2}{f_{t1}} + \frac{(f_{02}-f_{t2})^2}{f_{t2}} + \dots + \frac{(f_{0k}-f_{tk})^2}{f_{tk}} = \sum \frac{(f_{0i}-f_{ti})^2}{f_{ti}} \quad (\text{I-83})$$

4. Pearson a démontré que la distribution de cette quantité est approximativement celle du Khi deux avec ν degrés de liberté, avec $\nu = k - 1 - r$ (I-84) (r : nombre de paramètres qui caractérisent complètement la distribution théorique (dans le cas de la loi normale $r= 2$)) ;

5. Une fois le χ^2 relatif à notre échantillon déterminé, on le compare au $\chi_{\nu,\alpha}^2$ donné par une table (Annexe III) pour un degré de liberté connu ν et une probabilité au dépassement α (seuil de signification fixé à l'avance). Deux cas peuvent se poser :

a) Si le χ^2 est plus petit que le $\chi_{\nu,\alpha}^2$; l'on accepte l'hypothèse que le phénomène étudié suit la distribution théorique choisie et que les écarts entre les fréquences observées et les fréquences théoriques ne sont pas significatifs ;

b) Si le χ^2 est plus grand que le $\chi_{\nu,\alpha}^2$; l'on rejette l'hypothèse que H_0 considérée car les écarts sont significatifs, ce qui veut dire que les données expérimentales suivant une loi autre que celle de notre hypothèse, et l'on essaye une autre loi d'ajustement.

1.5.3. Test de Kolmogorov -Smirnov

C'est un test d'ajustement qui permet de comparer une distribution de valeurs observées à une distribution théorique. Ce test joue le même rôle que celui du Khi deux.

Pour accepter ou rejeter l'hypothèse H_0 que la loi choisie s'ajuste bien à notre échantillon, on considère la grandeur D que caractérise la différence entre la répartition empirique et la répartition théorique. La grandeur D peut être choisie de plusieurs façons. Cette grandeur D est elle-même une variable aléatoire dont la loi de répartition dans certains cas, pour N suffisamment grand, ne dépend pratiquement pas de la fonction $F(x)$.

Avec le test de Kolmogorov-Smirnov, on cherche la valeur maximale de la valeur absolue de la différence entre la fonction de répartition empirique $F_N(x)$ d'un échantillon de N valeurs et la fonction de répartition théorique $F(x)$ correspondante soit (*Sari, 2002a*):

$$D_N = D_{\max} = \max|F_N(x) - F(x)| \quad (\text{I-85}).$$

Kolmogorov a montré que, quelle que soit la fonction de répartition $F(x)$ d'une variable continue x lorsque le nombre d'observations augmente, la fonction de répartition de la grandeur $D_N\sqrt{N}$ tend vers : $\Pr(D_N\sqrt{N}) \rightarrow K(y) = \sum_{-\infty}^{+\infty} (-1)^k e^{-2k^2 y^2}$. (I-86).

Les valeurs de cette probabilité ont été tabulées (Annexe IV). On rejettera l'hypothèse que la loi choisie représente notre échantillon, au niveau de signification choisi, lorsque D_N est supérieure ou égal à d_n qui est la valeur de l'écart théorique.

I.6. Mise en application

Exercice 1.

Soit les températures moyennes annuelles T (en °C) données dans le tableau 1. On adopte le regroupement des données en classes de longueurs 0,5 °C, la borne inférieure minimale et 15,5 °C et la borne supérieure maximale 19 °C.

Tableau.

Année	T en °C	Année	T en °C	Année	T en °C	Année	T en °C
1955/1956	18,00	1965/1966	17,60	1975/1976	16,20	1985/1986	17,32
1956/1957	16,40	1966/1967	16,90	1976/1977	16,00	1986/1987	16,83
1957/1958	17,38	1967/1968	17,00	1977/1978	16,50	1987/1988	17,90
1958/1959	17,00	1968/1969	17,30	1978/1979	16,55	1988/1989	18,90
1959/1960	17,60	1969/1970	17,90	1979/1980	16,80	1989/1990	18,22
1960/1961	17,85	1970/1971	16,95	1980/1981	17,00	1990/1991	18,80
1961/1962	18,30	1971/1972	16,15	1981/1982	17,45	1991/1992	17,03
1962/1963	18,10	1972/1973	15,70	1982/1983	17,11		
1963/1964	17,40	1973/1974	16,40	1983/1984	17,95		
1964/1965	17,10	1974/1975	16,20	1984/1985	16,42		

- 1- Dresser le tableau des effectifs.
- 2- Dresser le tableau des fréquences cumulées croissantes et décroissantes.
- 3- Tracer la courbe des fréquences cumulées croissantes et décroissantes.
- 4- Déterminer graphiquement le point d'intersection entre ces deux courbes, quelles sont les coordonnées de ce point ? Que représente l'ordonnée de ce point ?

Solution.

- 1- Tableau des effectifs.

Tableau I.1. Effectifs des températures moyennes annuelles (°C) de la station de Béni-Bahdel.

T en °C	n_i	T en °C	n_i	T en °C	n_i	T en °C	n_i
15,70	1	16,90	1	17,45	1	18,90	1
16,00	1	16,95	1	17,60	2		
16,15	1	17,00	3	17,85	1		
16,20	2	17,03	1	17,90	2		
16,40	2	17,10	1	17,95	1		
16,42	1	17,11	1	18,00	1		
16,50	1	17,30	1	18,10	1		
16,55	1	17,32	1	18,22	1		
16,80	1	17,38	1	18,30	1		
16,83	1	17,40	1	18,80	1		

Ainsi, le regroupement des données en classes de longueurs 0,5 C°, en prenant la borne inférieure minimale égale à 15,5 °C et la borne supérieure maximale 19 C° se résume comme suite (tableau I.2) :

Tableau I.2. Fréquences absolues et relatives des températures moyennes annuelles (°C) de la station de Béni-Bahdel.

N°	T (°C)	n _i	f _i
1	15,5 ≤ T < 16	1	0,027
2	16 ≤ T < 16,5	7	0,189
3	16,5 ≤ T < 17	6	0,162
4	17 ≤ T < 17,5	11	0,297
5	17,5 ≤ T < 18	6	0,162
6	18 ≤ T < 18,5	4	0,108
7	18,5 ≤ T < 19	2	0,054

2- Tableau des fréquences cumulées croissantes et décroissantes (tableau I.3).

Tableau I.3. Fréquences absolues et relatives cumulées des températures moyennes annuelles (°C) de la station de Béni-Bahdel.

T (°C)	Effectif cumul croissant	Fréquence cumulé croissant	T (°C)	Effectif cumul décroissant	Fréquence cumulé décroissant
< 15,5	0	0	> 15,5	37	1
< 16	1	0,027	> 16	36	0,973
< 16,5	8	0,216	> 16,5	29	0,784
< 17	14	0,378	> 17	23	0,622
< 17,5	25	0,675	> 17,5	12	0,324
< 18	31	0,837	> 18	6	0,162
< 18,5	35	0,945	> 18,5	2	0,054
< 19	37	1	> 19	0	0

3- Courbe des fréquences cumulées croissantes et décroissantes (Fig. I.10).

4- Les coordonnées de l'intersection entre les deux courbes.
Que représente l'ordonnée de cette intersection ?

D'après la Fig. I.10, les coordonnées de l'intersection des courbes cumulées croissante et décroissante sont (17,20- 0,5).

L'ordonnée de ce point représente la *médiane* de l'échantillon.

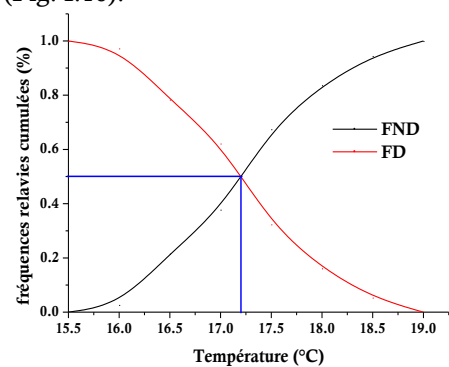


Fig. I.10. Fréquences cumulées croissantes et décroissantes des températures moyennes annuelles (°C) de la station de Béni-Bahdel.

Exercice 2.

Soit une série de précipitations annuelles enregistrées à une station pluviométrique dans le bassin versant de l'Oued Tafna :

Tableau.

Année	P (mm)	Année	P (mm)	Année	P (mm)
1955/1956	504,8	1965/1966	276,4	1975/1976	387,7
1956/1957	513	1966/1967	450,6	1976/1977	438
1957/1958	499	1967/1968	756,1	1977/1978	437,9
1958/1959	435	1968/1969	495,9	1978/1979	301,1
1959/1960	519	1969/1970	507,8	1979/1980	508,7
1960/1961	372,1	1970/1971	527,5		
1961/1962	607,5	1971/1972	599,8		
1962/1963	529,9	1972/1973	661,4		
1963/1964	613,6	1973/1974	729,1		
1964/1965	680,3	1974/1975	659,2		

- 1- Déterminer la moyenne arithmétique annuelle, la moyenne géométrique et la moyenne harmonique de l'échantillon.
- 2- Construire l'histogramme et la courbe des fréquences cumulées en prenant un nombre de classes égal à 6.
- 3- Déterminer la médiane, le mode, l'écart-type et le coefficient de variation.

Solution.

- 1- Calcul des moyennes.

Nous commençons par établir le tableau I.4 suivant. Ce tableau regroupe les pluies mesurées, leurs logarithmes népériens et leurs inverses.

Tableau I.4. Pluies mesurées, leurs logarithmes népériens et leurs inverses.

Années	Pluies mesurées (mm)	Ln P _i	$\frac{1}{P_i}$	Années	Pluies mesurées (mm)	Ln P _i	$\frac{1}{P_i}$
1955/1956	504,8	6,22	0,00198	1968/1969	495,9	6,21	0,00202
1956/1957	513	6,24	0,00195	1969/1970	507,8	6,23	0,00197
1957/1958	499	6,21	0,00200	1970/1971	527,5	6,27	0,00190
1958/1959	435	6,08	0,00230	1971/1972	599,8	6,40	0,00167
1959/1960	519	6,25	0,00193	1972/1973	661,4	6,49	0,00151
1960/1961	372,1	5,92	0,00269	1973/1974	729,1	6,59	0,00137
1961/1962	607,5	6,41	0,00165	1974/1975	659,2	6,49	0,00152
1962/1963	529,9	6,27	0,00189	1975/1976	387,7	5,96	0,00258
1963/1964	613,6	6,42	0,00163	1976/1977	438	6,08	0,00228
1964/1965	680,3	6,52	0,00147	1977/1978	437,9	6,08	0,00228
1965/1966	276,4	5,62	0,00362	1978/1979	301,1	5,71	0,00332
1966/1967	450,6	6,11	0,00222	1979/1980	508,7	6,23	0,00197
1967/1968	756,1	6,63	0,00132	Somme=	13011,4	155,64	0,05104

$$\bar{P} = \frac{1}{N} \sum_{i=1}^N P_i \Rightarrow \bar{P} = \frac{13011,4}{25} \Rightarrow \bar{P} = 520,456 \text{ mm}$$

$$P_G = \left(\prod_{i=1}^N P_i \right)^{\frac{1}{N}} \Rightarrow \ln P_G = \frac{1}{N} \sum_{i=1}^N \ln P_i \Rightarrow \ln P_G = \frac{155,64}{25} = 6,2256 \text{ et } P_G = e^{6,2256} \Rightarrow P_G = 505,53 \text{ mm}$$

$$P_H = \frac{N}{\sum_{i=1}^N \frac{1}{P_i}} \Rightarrow P_H = \frac{25}{0,05104} \Rightarrow P_H = 489,81 \text{ mm}$$

En effet : $P_H = 489,81 \text{ mm} < P_G = 505,53 \text{ mm} < \bar{P} = 520,456 \text{ mm}$

2- Histogramme et courbes des fréquences cumulées.

Nous commençons par ordonner la série de pluies par ordre croissant (ou décroissant). Le résultat est représenté dans le tableau I.5.

Tableau I.5. *Pluies mesurées, leurs logarithmes népériens et leurs inverses.*

P (mm)	n _i	P (mm)	n _i	P (mm)	n _i
276,4	1	504,8	1	659,2	1
301,1	1	507,8	1	661,4	1
372,1	1	508,7	1	680,3	1
387,7	1	513	1	729,1	1
435	1	519	1	756,1	1
437,9	1	527,5	1		
438	1	529,9	1		
450,6	1	599,8	1		
495,9	1	607,5	1		
499	1	613,6	1		

- Tracé de l'histogramme

Selon l'énoncé de l'exercice, l'échantillon est divisé en 6 classes égales.

Pour avoir des classes de pas égaux, nous utiliserons la formule suivante : $\text{Pas} = \frac{P_{\max} - P_{\min}}{K}$, avec K est le nombre de classe.

Dans notre cas : $\text{Pas} = \frac{756,1 - 276,4}{6} = 79,95 \text{ mm} \cong 80 \text{ mm}$.

On calcule ensuite pour chaque classe : son centre, sa fréquence absolue et relative (tableau I.6).

Tableau I.6. *Calcul des fréquences absolues et relatives d'une station pluviométrique (Bassin versant de l'Oued Tafna).*

N°	Classes	Centre de la classe	Effectif ou Fréquence absolue	Fréquence relative
1	276,4 ≤ P < 356,4	316,4	2	0,08
2	356,4 ≤ P < 436,4	396,4	3	0,12
3	436,4 ≤ P < 516,4	476,4	9	0,36
4	516,4 ≤ P < 596,4	556,4	3	0,12
5	596,4 ≤ P < 676,4	636,4	5	0,2
6	676,4 ≤ P < 756,4	716,4	3	0,12
Somme =			25	1

L'histogramme (Fig. I.11) est tracé en portant en abscisse les classes et en ordonnée soient les fréquences relatives soient les fréquences absolues. Les extrémités des bases des rectangles sont égales aux limites des classes.

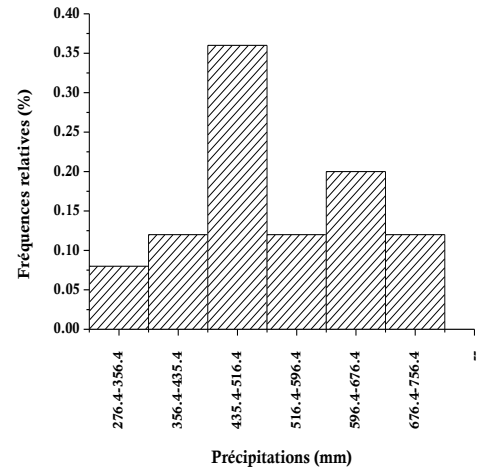


Fig. I.11. Histogramme des fréquences relatives des précipitations annuelles d'une station pluviométrique (Bassin versant de l'Oued Tafna).

- Calcul des fréquences cumulées (tableau I.7).

Tableau I.7. Courbes des fréquences cumulées croissantes et décroissantes des totaux pluviométriques annuels d'une station pluviométrique (Bassin versant de l'Oued Tafna).

Pluies (mm)	Effectif cumulé	Fréquence cumulée au non-dépassement (FND)	Pluie (mm)	Effectif cumulé	Fréquence cumulée au dépassement (FD)
< 276,4	0	0	> 276,4	25	1
< 356,4	2	0,08	> 356,4	23	0,92
< 436,4	5	0,2	> 436,4	20	0,8
< 516,4	14	0,56	> 516,4	11	0,44
< 596,4	17	0,68	> 596,4	8	0,32
< 676,4	22	0,88	> 676,4	3	0,12
< 756,4	25	1	> 756,4	0	0

Les courbes des fréquences cumulées croissantes et décroissantes des totaux pluviométriques annuels d'une station pluviométrique (bassin versant de l'Oued Tafna) sont représentées dans la Fig. I.12.

La somme des fréquences de toutes les valeurs plus petites que la limite supérieure d'un intervalle est appelée fréquence cumulée au non dépassement (FND). Ainsi **8%** des pluies annuelles considérées sont *inférieures* à **356,4 mm**. En outre, la somme des fréquences de toutes les valeurs plus grandes que la limite inférieure d'un intervalle est appelée fréquence cumulée au dépassement (FD). Ainsi, **92%** des pluies annuelles de notre échantillon sont *supérieures* à **356,4 mm**. On constate que : $FND + FD = 8\% + 92\% = 100\%$

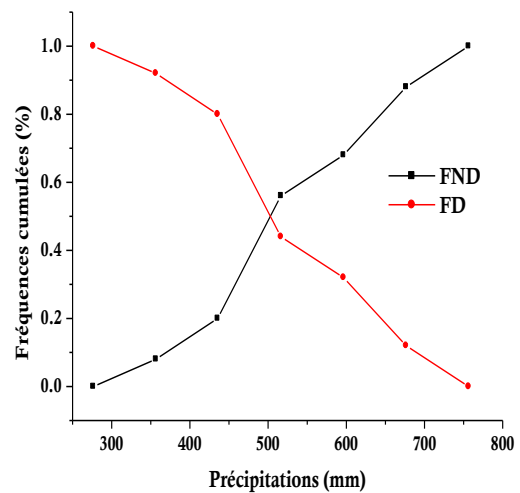


Fig. I.12. Courbes des fréquences cumulées croissantes et décroissantes des totaux pluviométriques annuels d'une station pluviométrique (Bassin versant de l'Oued Tafna).

3- Calcul de la médiane M_e , du mode M_o , de l'écart type s et du coefficient de variation CV.

- *Calcul de la médiane*

La médiane est la valeur qui partage l'échantillon en 50% supérieur et 50% inférieur. On peut l'obtenir soit graphiquement ou analytiquement.

i. *Graphiquement*

Le point d'intersection entre la courbe des fréquences cumulées croissantes et décroissantes, nous donne en ordonnée la fréquence 0,5 et en abscisse la valeur de la médiane.

Dans notre cas : $M_e=502,76$ mm (Fig. I.13).

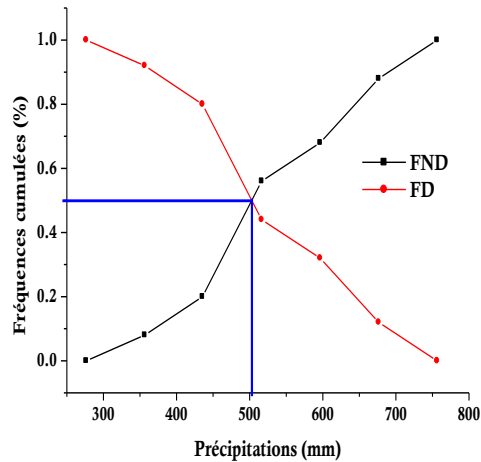


Fig. I.13. Détermination de la médiane graphiquement.

ii. *Analytiquement*

La formule est la suivante :

$$\frac{M_e - a_i}{a_{i+1} - a_i} = \frac{0,5 - f_i}{f_{i+1} - f_i}$$

Avec ;

a_i et a_{i+1} représente les limites de l'intervalle de la classe médianale; f_i et f_{i+1} leurs fréquences respectives.

Pour calculer la médiane, nous allons utiliser les valeurs des fréquences cumulées croissantes du tableau I.7.

Dans notre cas, la classe médianale est $436,4 \leq P < 516,4$ ainsi :

$$a_i = 436,4 \text{ mm} \Rightarrow f_i = 0,2 \text{ et } a_{i+1} = 516,4 \text{ mm} \Rightarrow f_{i+1} = 0,56$$

$$AN : \frac{M_e - 436,4}{516,4 - 436,4} = \frac{0,5 - 0,2}{0,56 - 0,2} \Rightarrow M_e = 503,07 \text{ mm}$$

- *Calcul du mode*

Le mode est la valeur la plus fréquente de l'échantillon. On peut l'obtenir graphiquement en utilisant l'histogramme des fréquences relatives (Fig. I.14) et en appliquant la formule

$$\text{suivante : } M_o = a_i + \left(\frac{\Delta_1}{\Delta_1 + \Delta_2} \right) \times (a_{i+1} - a_i)$$

D'après l'histogramme, la classe modale est

$$436,4 \leq P < 516,4$$

$$a_i = 436,4 \text{ mm ; } a_{i+1} = 516,4 \text{ mm}$$

$$\Delta_1 = \Delta_2 = 0,36 - 0,12 = 0,24$$

$$M_o = 436,4 + 80 \times \left(\frac{0,24}{0,48} \right) \Rightarrow M_o = 476,4 \text{ mm}$$

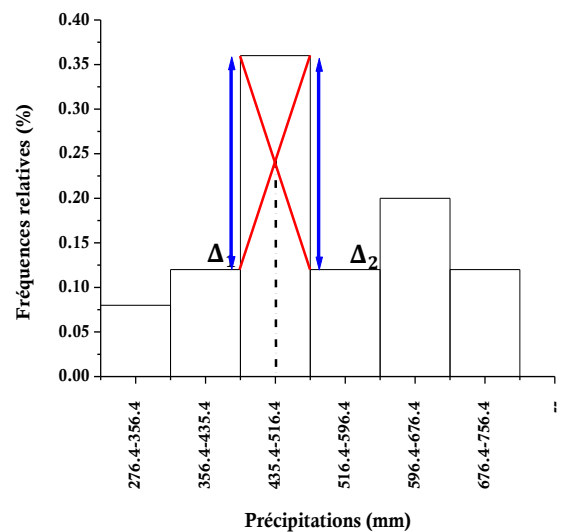


Fig. I.14. Détermination de la médiane graphiquement

- Calcul de l'écart type

L'écart type est calculé grâce à l'une des formules suivantes :

$$S_p = \sqrt{\frac{\sum(P_i - \bar{P})^2}{n-1}}; \quad S_p = \sqrt{\frac{\sum P_i^2 - n\bar{P}^2}{n-1}} \quad \text{ou} \quad S_p = \sqrt{\frac{\sum P_i^2 - \frac{(\sum P_i)^2}{n}}{n-1}}$$

Pour pouvoir utiliser ces formules, on établit le tableau I.8 suivant :

Tableau I.8. Détermination de la médiane graphiquement.

Années	Pluies (mm)	$(P_i - \bar{P})^2$	P_i^2	Années	Pluies (mm)	$(P_i - \bar{P})^2$	P_i^2
1955/1956	504,8	245,110336	254823,04	1968/1969	495,9	602,997136	245916,81
1956/1957	513	55,591936	263169	1969/1970	507,8	160,174336	257860,84
1957/1958	499	460,359936	249001	1970/1971	527,5	49,617936	278256,25
1958/1959	435	7302,72794	189225	1971/1972	599,8	6295,47034	359760,04
1959/1960	519	2,119936	269361	1972/1973	661,4	19865,2111	437449,96
1960/1961	372,1	22009,5027	138458,41	1973/1974	729,1	43532,3187	531586,81
1961/1962	607,5	7576,65794	369056,25	1974/1975	659,2	19249,8975	434544,64
1962/1963	529,9	89,189136	280794,01	1975/1976	387,7	17624,1555	150311,29
1963/1964	613,6	8675,80474	376504,96	1976/1977	438	6798,99194	191844
1964/1965	680,3	25550,1043	462808,09	1977/1978	437,9	6815,49314	191756,41
1965/1966	276,4	59563,3311	76396,96	1978/1979	301,1	48117,0547	90661,21
1966/1967	450,6	4879,86074	203040,36	1979/1980	508,7	138,203536	258775,69
1967/1968	756,1	55528,0947	571687,21	Somme=	13011,4	361188,042	7133049,24

$$S_p = \sqrt{\frac{\sum(P_i - \bar{P})^2}{n-1}} = \sqrt{\frac{361188,042}{24}} \Rightarrow S_p = 122,676 \text{ mm} ;$$

$$S_p = \sqrt{\frac{\sum P_i^2 - n\bar{P}^2}{n-1}} = \sqrt{\frac{7133049,24 - 24 \times 520,456^2}{24}} \Rightarrow S_p = 122,676 \text{ mm} ;$$

$$S_p = \sqrt{\frac{\sum P_i^2 - \frac{(\sum P_i)^2}{n}}{n-1}} = \sqrt{\frac{7133049,24 - \frac{(13011,4)^2}{25}}{24}} \Rightarrow S_p = 122,676 \text{ mm}.$$

- Calcul du coefficient de variation

$$CV = \frac{S_p}{\bar{P}} = \frac{122,676}{520,456} \Rightarrow CV = 0,2357 \approx 0,24$$

Exercice 3.

Soit une série de précipitations annuelles de 24 ans enregistrées à une station pluviométrique dans le bassin versant de l'Oued FODDA.

Tableau.

P (mm)	n_i	f_i	$f_{i \text{ cum}} \nearrow$
[245, 300[0,13	
[300, 355[0,33	
[355, 410[0,04	
[410, 465[0,71
[465, 520[0,75
[520, 575[0,79
[575, 630[0,92
[630, 685[1,00

- 1- Compléter le tableau statistique ci-dessus.
- 2- Représenter le polygone des fréquences et calculer le mode.
- 3- Tracer la courbe des fréquences cumulées croissantes et décroissantes
- 4- Quelle est la portion des précipitations qui ont : au moins de 410 mm ? moins de 465 mm ? entre 350 mm et 650 mm (détermination graphiquement).

Solution.

1- Compléter le tableau

Nous commençons par le calcul des fréquences cumulées croissantes. Ensuite, nous calculons les fréquences relatives f_i , avec $f_i = f_{\text{cum } i} - f_{\text{cum } i-1}$. Par exemple : $f_4 = f_{\text{cum } 4} - f_{\text{cum } 3} = 0,71 - 0,50 = 0,21$

Le calcul des fréquences absolues n_i est effectué en utilisant la formule suivante : $f_i = \frac{n_i}{N} \Rightarrow n_i = f_i \times N$

Avec, $N=24$.

Le résultat du calcul doit être arrondi pour obtenir un chiffre entier naturel pour représenter la fréquence absolue ou l'effectif de chaque classe (Tableau I.9).

Tableau I.9. Calcul des fréquences absolues, relatives et cumulées croissantes d'une station pluviométrique (Bassin versant de l'Oued Fodda).

N°	P (mm)	Centre de la classe	n_i	f_i	$f_{i \text{ cum}} \nearrow$
1	[245, 300[272,5	3	0,13	0,13
2	[300, 355[375,5	8	0,33	0,46
3	[355, 410[382,5	1	0,04	0,50
4	[410, 465[437,5	5	0,21	0,71
5	[465, 520[492,5	1	0,04	0,75
6	[520, 575[547,5	1	0,04	0,79
7	[575, 630[602,5	3	0,13	0,92
8	[630, 685[657,5	2	0,08	1,00
Somme=			24	1	

2- Tracer le polygone des fréquences et calculer le mode

- *Polygone des fréquences*

Pour tracer le polygone des fréquences (Fig. I.15), il faut tracer l'histogramme des fréquences relatives (ou absolue) du tableau I.9.

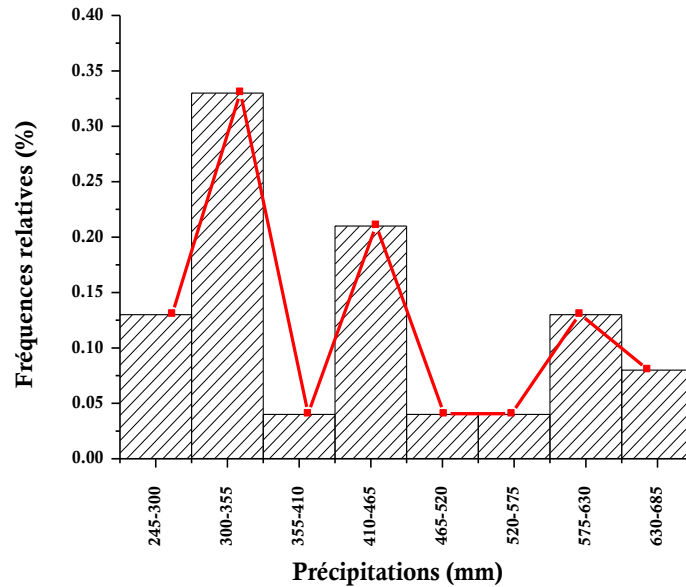


Fig. I.15. Histogramme et polygone des fréquences relatives des totaux pluviométriques annuels d'une station pluviométrique (Bassin versant d'Oued Fodda).

- *Calcul du mode*

$$M_o = a_i + \left(\frac{\Delta_1}{\Delta_1 + \Delta_2} \right) \times (a_{i+1} - a_i)$$

D'après l'histogramme, la classe modale est [300, 355[, ainsi :

$$a_i = 300 \text{ mm} ; a_{i+1} = 355 \text{ m} ;$$

$$\Delta_1 = 0,33 - 0,13 = 0,20$$

$$\text{et } \Delta_2 = 0,33 - 0,04 = 0,29$$

$$\text{AN : } M_o = 300 + 55 \times \left(\frac{0,20}{0,49} \right) \Rightarrow M_o = 322,44 \text{ mm}$$

3- Courbe des fréquences cumulées croissantes et décroissantes (tableau I.10 et Fig. I.16)

Tableau I.10. Calcul des fréquences cumulées croissantes et décroissantes d'une station pluviométrique (Bassin versant de l'Oued Fodda).

N°	P (mm)	f_i	$f_{i\text{ cum } \nearrow}$	$f_{i\text{ cum } \searrow}$
1	[245,300[0,13	0,13	1,00
2	[300,355[0,33	0,46	0,97
3	[355,410[0,04	0,50	0,54
4	[410,465[0,21	0,71	0,50
5	[465,520[0,04	0,75	0,29
6	[520,575[0,04	0,79	0,25
7	[575,630[0,13	0,92	0,21
8	[630,685[0,08	1,00	0,08

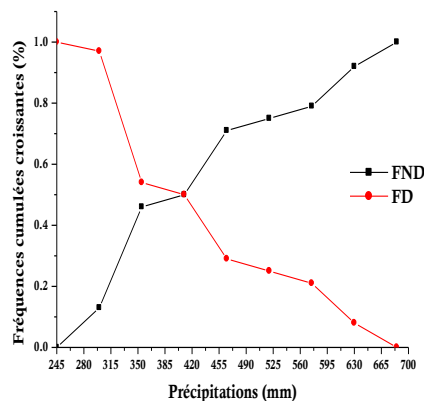


Fig. I.16. Courbes des fréquences cumulées croissantes et décroissantes d'une station pluviométrique (Bassin versant de l'Oued Fodda).

4- La portion des précipitations qui ont : au moins de 410 mm ? moins de 465 mm ? entre 350 mm et 650 mm (détermination graphique).

- **Au moins de 410 mm :** Pour trouver une probabilité d'avoir une précipitation au moins de 410 mm, autrement dit, supérieure à 410 mm, on procède comme suite : on cherche la valeur 410 mm sur l'axe des abscisses. Une projection verticale sera réalisée à partir de cette valeur sur la courbe FD. A partir de ce point d'intersection obtenue, on procède à une projection horizontale vers l'axe des ordonnées (Fig. I.17). La valeur trouvée est de 0,50. Autrement dit, qu'il y a **50% de chance d'avoir des précipitations supérieures de 410 mm**. Dans ce cas la valeur 410 mm représente la médiane de la série des totaux pluviométriques annuels de cette station.

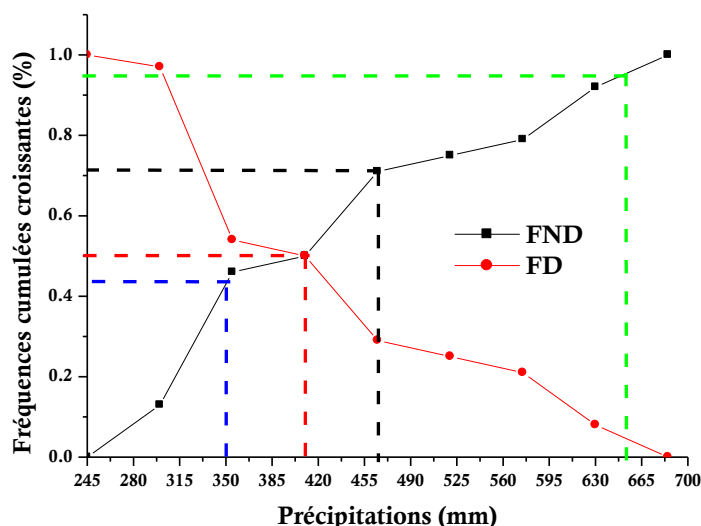


Fig. I.17. Portion des précipitations. Au moins de 410 mm (ligne rouge). Moins de 465 mm (ligne noire), entre 350 mm et 650 mm (lignes verte et bleu).

- **Moins de 465 mm :** Pour trouver une probabilité d'avoir une précipitation moins de 465mm, autrement dit, inférieur de 465 mm, on cherche la valeur 465 mm sur l'axe des abscisses. Une projection verticale sera réalisée à partir de cette valeur sur la courbe FND. A partir de ce point d'intersection obtenue, on procède à une projection horizontale vers l'axe des ordonnées (Fig. I.17). La valeur trouvée est de 0,71 qui veut dire que qu'il y a **71% de chance d'avoir des précipitations inférieures de 465 mm**.

- **Entre 350 mm et 650 mm :** De la même façon, on trouve (Fig. I.17) :

Moins de 350 mm \Rightarrow 0,44 et Moins de 650 mm \Rightarrow 0,95 Entre 350 mm et 650 mm : $0,95 - 0,44 = 0,43$.

Ainsi ; on a **43% de chance d'avoir des précipitations entre 350 mm et 650 mm**.

Exercice 4.

Une série de 60 pluies annuelles suivant une loi de distribution normale de moyenne de 680 mm et d'écart-type de 30 mm :

- 1- Quelle est la probabilité des pluies supérieures à 700 mm ;
- 2- Quelle est la probabilité des pluies inférieures à 610 mm ;
- 3- Quelle est la probabilité et le nombre d'années des pluies comprises entre 610 et 700 mm.

Anonyme / USTHB – FGC – IMIE-1112/ TD N°1 d'Hydrologie.

Solution.

Comme la série de 60 pluies annuelles **s'ajuste à une loi normale**, on doit utiliser, dans ce cas, la table de Gauss (Annexe I) pour calculer les différentes probabilités demandées.

Pour pouvoir utiliser la table de Gauss, on doit d'abord calculer les variables réduites correspondants aux pluies considérées (700 mm et 610 mm).

$$u_{700} = \frac{P - \bar{P}}{S_p} = \frac{700 - 680}{30} \Rightarrow u_{700} = 0,67$$

$$\text{et } u_{610} = \frac{P - \bar{P}}{S_p} = \frac{610 - 680}{30} \Rightarrow u_{610} = -2,33$$

- 1- Probabilité d'avoir une pluie supérieure à 700 mm

La table de Gauss donne les fréquences au non dépassement (FND). Pour $u_{700} = 0,67 \Rightarrow \text{FND} = 0,74857$, mais comme on cherche la probabilité de dépassement c'est-à-dire FD, on utilise, dans ce cas, la relation $\text{FD} + \text{FND} = 1$ pour trouver FD. AN : $\text{FD} = 1 - \text{FND} = 1 - 0,74857 = 0,2514$ ou 25,14%. Ainsi, *Il y a 25,14 % de chance d'avoir une pluie supérieure à 700 mm.*

- 2- Probabilité d'avoir une pluie inférieure à 621 mm

La table de Gauss donnant les fréquences au non dépassement (FND) donne pour $u_{610} = -2,33 \Rightarrow \text{FND} = (1 - 0,99010) = 0,0099$ ou 0,99%. Ainsi, *Il y a 0,99 % de chance d'avoir une pluie inférieure à 680 mm.*

- 3- Probabilité et nombre d'années des pluies comprises entre 680 et 700mm.

La Fig. I.18 illustre bien les différentes probabilités. Les surfaces sous la courbe de densité de probabilité correspondent aux différentes probabilités. Ainsi $S1 = 0,2514$ correspond à la probabilité d'avoir $P > 700$ mm ; $S2 = 0,0099$ correspond à la probabilité d'avoir $P < 680$ mm ; et $S3$ correspond à la probabilité d'avoir $681 \text{ mm} < P < 700$ mm.

La surface sous la courbe densité de probabilité est égale à 1 par définition. Donc, $S3 = 1 - S1 - S2$
 $S3 = 1 - 0,25414 - 0,0099 \Rightarrow S3 = 0,73549$ ou 73,55%.

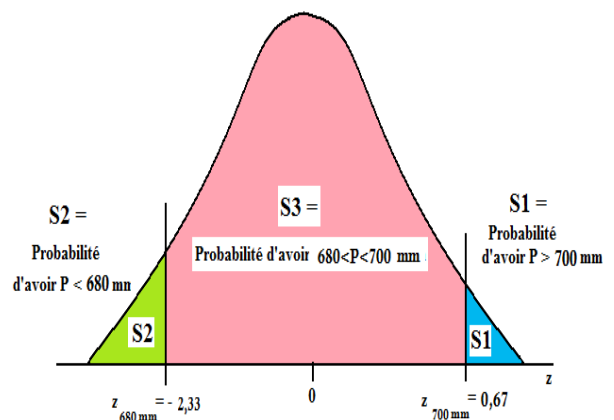


Fig. I.18. Les différentes probabilités comprises entre 680 mm et 700 mm.

Comme notre échantillon compte 60 années le nombre d'années où on aura $680 \text{ mm} < P < 700 \text{ mm}$ est égal à $60 \times 0,73549 = 44 \text{ ans}$.

Exercice 5.

Soit une série de pluies annuelles recueillies à une station pluviométrique sur Oued FODDA.

Tableau.

Année	Pluies (mm)	Année	Pluies (mm)	Année	Pluies (mm)
1969	456	1977	321	1985	315
1970	349	1978	449	1986	245
1971	622	1979	653	1987	342
1972	520	1980	394	1988	297
1973	494	1981	587	1989	302
1974	302	1982	681	1990	321
1975	448	1983	257	1991	411
1976	576	1984	347	1992	443

- 1- Déterminer la moyenne, l'écart type et le coefficient de variation de l'échantillon ;
- 2- Ajuster une loi de Gauss à la série des pluies annuelles ;
- 3- Vérifier la validité de l'ajustement à l'aide du test du χ^2 à 95% ;
- 4- Déterminer la pluie annuelle de période de retour de 25 ans ainsi que son intervalle de confiance à 90% ;
- 5- Déterminer les périodes de retour théorique et expérimentale de la pluie de 622,6 mm
- 6- Calculer les intervalles de confiance de la moyenne et de l'écart type à 80% ;
- 7- Tracer la courbe enveloppe de l'intervalle de confiance à un seuil de confiance de 95%.

Anonyme / USTHB – FGC – IMIE-1112/ TD N°4 d'Hydrologie

Solution.

- 1- Calcul de la moyenne, l'écart type et le coefficient de variation de l'échantillon

On établit le tableau I.11 en vue de faciliter le calcul de $\sum_{i=1}^n P_i$ et $\sum_{i=1}^n (P_i - \bar{P})^2$.

Tableau I.11. Pluies mesurées, leurs carrées et l'écart à la moyenne carrée.

P_i	$(P_i - \bar{P})$	$(P_i - \bar{P})^2$	P_i	$(P_i - \bar{P})$	$(P_i - \bar{P})^2$
456	33,83	1144,469	587	164,83	27168,93
349	-73,17	5353,849	681	258,83	66992,97
622	199,83	39932,03	257	-165,17	27281,13
520	97,83	9570,709	347	-75,17	5650,529
494	71,83	5159,549	315	-107,17	11485,41
302	-120,17	14440,83	245	-177,17	31389,21
448	25,83	667,1889	342	-80,17	6427,229
576	153,83	23663,67	297	-125,17	15667,53
321	-101,17	10235,37	302	-120,17	14440,83
449	26,83	719,8489	321	-101,17	10235,37
653	230,83	53282,49	411	-11,17	124,7689
394	-28,17	793,5489	443	20,83	433,8889
		Somme	10132	382261,333	

$$\bar{P} = \frac{1}{N} \sum_{i=1}^N P_i \quad \text{AN. } \bar{P} = \frac{10132}{24} \Rightarrow \bar{P} = 422,17 \text{ mm}$$

$$S_p = \sqrt{\frac{\sum_{i=1}^N (P_i - \bar{P})^2}{N-1}} \text{ AN. } S_p = \sqrt{\frac{382261,3}{24-1}} = \sqrt{16620,058} \Rightarrow S_p = 128,92 \text{ mm}$$

$$Cv = \frac{S_p}{\bar{P}} \text{ AN. } Cv = \frac{128,92}{422,17} \Rightarrow Cv = 0,30537$$

2- Ajustement d'une loi normale à l'échantillon.

Le tableau I.12 indique les étapes de calcul pour ajuster une loi normale à notre échantillon.

- *Colonne. 1* : Les valeurs de l'échantillon, dans notre cas se sont les pluies annuelles mesurées à la station pluviométrique sur Oued FODDA ;
- *Colonne. 2* : Pluies annuelles classées ;
- *Colonne. 3* : Numéro d'ordre n_i des valeurs classées ;
- *Colonne. 4* : Fréquences expérimentales au non-dépassement calculé par la formule de Hazen $FND = \frac{n_i - 0,5}{N}$;
- *Colonne. 5* : Variables centrées réduites expérimentale u_i c'est-à-dire celles qu'on a tiré de la table de Gauss et qui correspondent aux FND de la colonne 4 ;
- *Colonne. 6* : Variables centrées réduites théoriques de Gauss $u_i = \frac{P_i - \bar{P}}{S_p}$;
- *Colonne. 7* : Pluies théoriques $P_i = \bar{P} + u_i \times S_p$.

Tableau I.12. Ajustement d'une loi normale à la série de pluies annuelles recueillies à une station pluviométrique (Oued FODDA).

1	2	3	4	5	6	7
Valeurs Mes.	Valeurs Clas.	Ordre	Fréq. Exp.	u_i Exp.	u_i Théo.	Valeurs Théo.
456	245	1	0,0208	-2,037	-1,529	225,05
349	257	2	0,0625	-1,534	-1,436	237,04
622	297	3	0,1042	-1,258	-1,126	277,01
520	302	4	0,1458	-1,054	-1,087	282,03
494	302	5	0,1875	-0,887	-1,087	282,03
302	315	6	0,2292	-0,742	-0,986	295,05
448	321	7	0,2708	-0,610	-0,940	300,99
576	321	8	0,3125	-0,489	-0,940	300,99
321	342	9	0,3542	-0,374	-0,777	322,00
449	347	10	0,3958	-0,264	-0,738	327,03
653	349	11	0,4375	-0,157	-0,723	328,96
394	394	12	0,4792	-0,052	-0,374	373,95
587	411	13	0,5208	0,052	-0,242	390,97
681	443	14	0,5625	0,157	0,007	423,07
257	448	15	0,6042	0,264	0,045	427,97
347	449	16	0,6458	0,374	0,053	429,00
315	456	17	0,6875	0,489	0,107	435,96
245	494	18	0,7292	0,610	0,402	474,00
342	520	19	0,7708	0,742	0,604	500,04
297	576	20	0,8125	0,887	1,038	555,99
302	587	21	0,8542	1,054	1,124	567,08
321	622	22	0,8958	1,258	1,395	602,01
411	653	23	0,9375	1,534	1,636	633,08
443	681	24	0,9792	2,037	1,853	661,06

Dans la première figure (Fig. I.19), on porte sur le papier millimétré (où les 2 échelles sont arithmétiques) en abscisses les variables réduites théorique de la colonne 6 et en ordonnées les pluies correspondantes de la colonne 2.

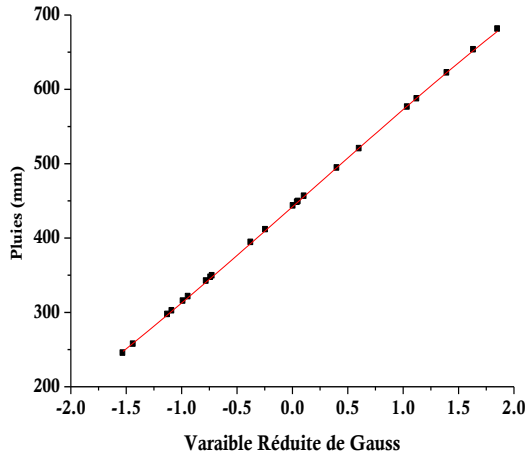


Fig. I.20. Ajustement d'une loi normale. Papier millimétré.

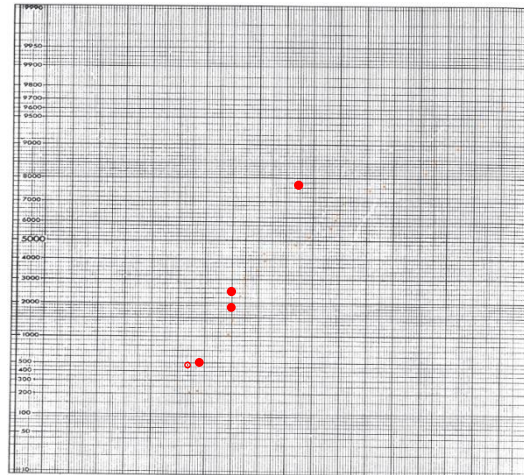


Fig. I.19. Ajustement d'une loi normale. Papier de probabilité de Gauss.

Dans la deuxième figure (Fig. I.20), on porte sur le papier de probabilité de Gauss, en abscisse les fréquences expérimentales (FND) de la colonne 4 et en ordonnée les pluies correspondantes de la colonne 2. Ce sont les points expérimentaux.

On peut tracer la droite de Henry soit sur du papier millimétré ou sur du papier de probabilité normale. La droite de Henry passe par deux points dont les coordonnées sont :

- *Sur le papier millimétré*
 Abscisses : les variables réduites u_1 et u_2 ;
 Ordonnées : les pluies correspondantes P_1 et P_2 .
- *Sur le papier de probabilité normale (papier de Gauss)*
 Abscisses : les fréquences F_1 et F_2 ;
 Ordonnées : les pluies correspondantes P_1 et P_2 .

On prend arbitrairement $F_1 = 0,5$ et $F_2 = 0,9$.

$$F_1 = 0,5 \rightarrow u_1 = 0 \quad \text{et} \quad P_1 = \bar{P} + u_1 \times S = 422,17 \text{ mm}$$

$$F_2 = 0,9 \rightarrow u_2 = 1,28 \quad \text{et} \quad P_2 = \bar{P} + u_2 \times S = 422,17 + 1,28 \times 128,9 = 587,16 \text{ mm}$$

3- Test du χ^2 à 95%.

Le tableau I.13 indique les étapes pour valider l'ajustement à l'aide du test du χ^2 à 95%.

- *Colonne. 1* : Numéro de la classe ;
- *Colonne. 2* : Borne inférieure des classes ;
- *Colonne. 3* : Borne supérieure des classes ;
- *Colonnes. 4 et 5* : Variable centrer réduite de Gauss : $u_i = \frac{P_i - \bar{P}}{S_P}$;

- *Colonnes. 6 et 7*: Probabilités au non-dépassement relatives aux bornes respectives, que l'on trouve en utilisant la table de Gauss ;
- *Colonne. 8*: Fréquence expérimentale de chaque classe. Dans ce cas, elle est égale à 6 ;
- *Colonne. 9*: Fréquence théorique de chaque classe $ft_i = N (F(u_2) - F(u_1))$;
- *Colonne. 10*: La valeur du $\chi_i^2 = \frac{(fo_i - ft_i)^2}{ft_i}$ pour chaque classe.

Tableau I.13. Application du test du χ^2 .

1	2	3	4	5	6	7	8	9	10
N°	X ₁	X ₂	u ₁	u ₂	F(u ₁)	F(u ₂)	fo _i	ft _i	χ ²
1	− ∞	320	− ∞	-0,792	0	0,214	6	5,14	0,145
2	320	400	-0,792	-0,172	0,214	0,432	6	5,22	0,115
3	400	500	-0,172	0,604	0,432	0,727	6	7,09	0,166
4	500	+∞	0,604	+ ∞	0,727	1	6	6,55	0,047
Somme							24	0,473	

La somme des $\chi^2 = 0,473$.

On cherche maintenant le $\chi_{v,\alpha}^2$, sachant que $\alpha = 0,95$ et $v =$ nombre de degrés de liberté $= k - 1 - r$

$v = 1$ où $k =$ nombre de classes $= 4$, $r = 2 =$ nombre de paramètres qui déterminent la loi normale.

La table du χ^2 (Annexe III) donne $\chi_{1,0,05}^2 = 3,84$.

Comme le χ^2 calculé est plus petit que celui donné par la table, on accepte donc, *l'hypothèse qu'une loi normale ayant une moyenne égale à 422,17 mm et un écart-type égal à 128,9 mm représente la population de pluies annuelles de la station pluviométrique (Oued FODDA).*

4- Détermination de la pluie annuelle de période de retour de 25 ans et de son intervalle de confiance à 90%.

- *Pluie annuelle de période de retour de 25 ans.*

$$T = 25 \Rightarrow FD = \frac{1}{25} = 0,04 \Rightarrow FND = 0,96 \Rightarrow u_{25} = 1,75$$

$$P_{25} = \bar{P} + u_{25} \times S_p = 422,17 + 1,75 \times 128,9 \Rightarrow P_{25} = 647,75 \text{ mm}$$

- *L'intervalle de confiance à 90% de pluie annuelle de période de retour de 25 ans*

IC à 90 % : L'intervalle de confiance est donné par la formule suivante :

$$P_{25} - u_{\frac{1-\alpha}{2}} \frac{S_p}{\sqrt{2n}} \sqrt{2 + u_p^2} < x_p < P_{25} + u_{\frac{1-\alpha}{2}} \frac{S_p}{\sqrt{2n}} \sqrt{2 + u_p^2}$$

$$\alpha = 0,90 \Rightarrow 1 - \alpha = 0,10 \Rightarrow \frac{1 - \alpha}{2} = 0,05 \text{ et } u_{\frac{1-\alpha}{2}} = 1,64$$

u_{25} calculé précédemment $u_{25} = 1,75$

Donc ;

$$647,75 - 1,64 \frac{128,9}{\sqrt{2 \times 24}} \sqrt{2 + (1,75)^2} < x_p < 647,75 + 1,64 \frac{128,9}{\sqrt{2 \times 24}} \sqrt{2 + (1,75)^2}$$

$$\Rightarrow 579,1 \text{ mm} < x_p < 716,4 \text{ mm}$$

5- Détermination des périodes de retour théorique et expérimentale de la pluie de 622,6 mm

- *Période de retour théorique*

Par définition la période de retour théorique est définie comme suit : $T = \frac{1}{FD} = \frac{1}{1-FND}$

La période de retour théorique est calculée à partir de la FND théorique qui correspond à la variable réduite de Gauss de la pluie de 622,6 mm.

$$u = \frac{P_i - \bar{P}}{s} = \frac{622,6 - 422,17}{128,9} \Rightarrow u = 1,55$$

Pour $u = 1,55$ la table de Gauss donne une FND égale à 0,93943 et une FD = $1 - 0,93943 = 0,06057$

Ainsi, la période de retour $T = \frac{1}{FD} = \frac{1}{0,06057} \Rightarrow T = 16,5 \text{ ans.}$

- *Période de retour expérimentale*

La période de retour expérimentale est égale à l'inverse de la FD expérimentale qui est égale à $1 - FND$ expérimentale.

$$\text{Selon la formule de Hazen } FND_{Exp} = \frac{m-0,5}{N} = \frac{22-0,5}{24} \Rightarrow FND_{Exp} = 0,986$$

$$FD_{Exp} = 1 - 0,986 \Rightarrow FD_{Exp} = 0,104$$

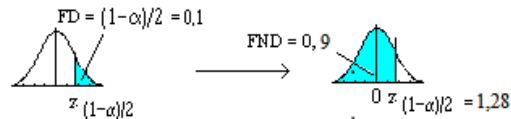
$$T = \frac{1}{0,104} \Rightarrow T = 9,6 \text{ ans.}$$

6- Calcul des intervalles de confiance de la moyenne et de l'écart type à 80%.

- *A : IC à 80% de la moyenne*

$$\bar{P} = 422,17 \text{ mm et } S_p = 128,9 \text{ mm}$$

$$\bar{P} - u_{\frac{1-\alpha}{2}} \frac{S_p}{\sqrt{n}} < \mu < \bar{P} + u_{\frac{1-\alpha}{2}} \frac{S_p}{\sqrt{n}}$$



FND = 0,9 d'où $u = 1,28$ et

$$422,17 - 1,28 \times \frac{128,9}{\sqrt{24}} < \mu < 422,17 + 1,28 \times \frac{128,9}{\sqrt{24}} \Rightarrow 455,8 \text{ mm} < \mu < 388,5 \text{ mm}$$

- *B : IC à 80% de l'écart-type*

$$S_p - u_{\frac{1-\alpha}{2}} \frac{S_p}{\sqrt{2n}} < \sigma < S_p + u_{\frac{1-\alpha}{2}} \frac{S_p}{\sqrt{2n}}$$

$$\Rightarrow 128,9 - 1,28 \times \frac{128,9}{\sqrt{2 \times 24}} < \sigma < 128,9 + 1,28 \times \frac{128,9}{\sqrt{2 \times 24}} \Rightarrow 152,7 \text{ mm} < \sigma < 105,1 \text{ mm}$$

7- Tracé de la courbe enveloppe de l'intervalle de confiance à un seuil de confiance de 95%.

On calcule les intervalles de confiances pour les pluies dont les variables réduites (ces variables sont choisis arbitrairement) sont : - 1, - 2, 0, 1 et 2. Ainsi, les pluies calculées sont comme suite :

$$P_1 = 422,17 - 2 \times 128,9 = 164,37 \text{ mm} ; P_2 = 422,17 - 1 \times 128,9 = 293,27 \text{ mm} ;$$

$$P_3 = 422,17 - 0 \times 128,9 = 422,17 \text{ mm} ; P_4 = 422,17 + 1 \times 128,9 = 551,07 \text{ mm} ;$$

$$P_5 = 422,17 + 2 \times 128,9 = 679,97 \text{ mm.}$$

$$\alpha = 0,95 \Rightarrow 1 - \alpha = 0,05 \Rightarrow \frac{1 - \alpha}{2} = 0,025 \text{ et } u_{\frac{1-\alpha}{2}} = 1,96$$

$$P_i - u_{\frac{1-\alpha}{2}} \frac{S_p}{\sqrt{2n}} \sqrt{2 + u_p^2} < \hat{P}_i < X_p + u_{\frac{1-\alpha}{2}} \frac{S_p}{\sqrt{2n}} \sqrt{2 + u_p^2}$$

$$164,37 - 1,96 \times \frac{128,9}{\sqrt{2 \times 24}} \times \sqrt{2 + (-2)^2} < \hat{P}_1 < 164,37 + 1,96 \times \frac{128,9}{\sqrt{2 \times 24}} \times \sqrt{2 + (-2)^2} \Rightarrow 253,7 \text{ mm} < \hat{P}_1 < 75,0 \text{ mm}$$

$$293,27 - 1,96 \times \frac{128,9}{\sqrt{2 \times 24}} \times \sqrt{2 + (-1)^2} < \hat{P}_2 < 293,27 + 1,96 \times \frac{128,9}{\sqrt{2 \times 24}} \times \sqrt{2 + (-1)^2} \Rightarrow 356,4 \text{ mm} < \hat{P}_2 < 230,1 \text{ mm}$$

$$422,17 - 1,96 \times \frac{128,9}{\sqrt{2 \times 24}} \times \sqrt{2 + (0)^2} < \hat{P}_3 < 422,17 + 1,96 \times \frac{128,9}{\sqrt{2 \times 24}} \times \sqrt{2 + (0)^2} \Rightarrow 473,7 \text{ mm} < \hat{P}_3 < 370,6 \text{ mm}$$

$$551,07 - 1,96 \times \frac{128,9}{\sqrt{2 \times 24}} \times \sqrt{2 + (1)^2} < \hat{P}_4 < 551,07 + 1,96 \times \frac{128,9}{\sqrt{2 \times 24}} \times \sqrt{2 + (1)^2} \Rightarrow 614,2 \text{ mm} < \hat{P}_4 < 487,9 \text{ mm}$$

$$679,97 - 1,96 \times \frac{128,9}{\sqrt{2 \times 24}} \times \sqrt{2 + (2)^2} < \hat{P}_5 < 679,97 + 1,96 \times \frac{128,9}{\sqrt{2 \times 24}} \times \sqrt{2 + (2)^2} \Rightarrow 769,3 \text{ mm} < \hat{P}_5 < 590,6 \text{ mm}$$

Le tracé de la courbe enveloppe de l'intervalle de confiance à un seuil de confiance de 95% est représenté dans la figure I.21.

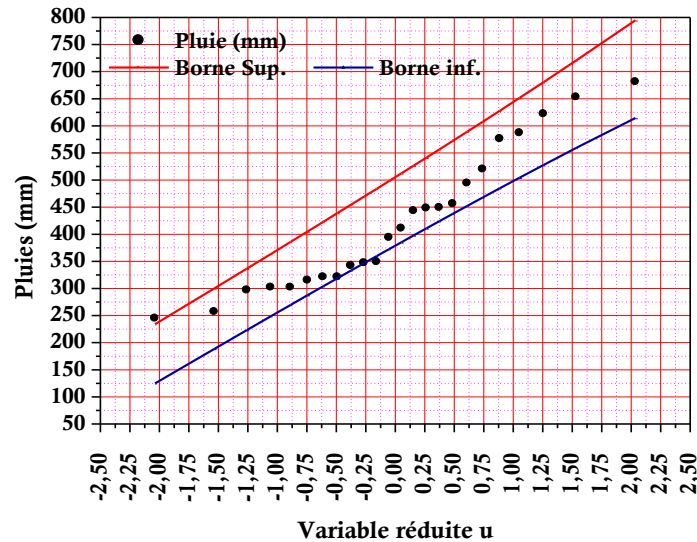


Fig. I.21. Courbe enveloppe de l'intervalle de confiance à un seuil de confiance de 95%.

Exercice 6.

- 1- Ajuster une loi Log-normale à l'échantillon de l'exercice précédent ;
- 2- Vérifier grâce au test du χ^2 au seuil de signification de 95% la validité de cet ajustement. Comparer avec le résultat précédent et conclure ;
- 4- Calculer l'intervalle de confiance de la moyenne à 80% ;
- 5- Calculer les pluies décennale et centennale ainsi que leurs intervalles de confiances à 80%.

Anonyme / USTHB – FGC – 1MIE-1112 / TD N°4 d'Hydrologie

Solution.

- 1- Ajustement d'une loi log-normale à l'échantillon.

Le tableau I.14 indique les étapes de calcul pour ajuster une loi log normale à notre échantillon.

- *Colonne. 1* : Valeurs de l'échantillon, dans notre cas, représentent les pluies annuelles mesurées à la station pluviométrique (Oued FODDA) ;
- *Colonne. 2* : Pluies annuelles classées ;

- Colonne. 3 : Numéro d'ordre « n_i » des valeurs classées ;
- Colonne. 4 : Fréquences expérimentales au non-dépassement calculé par la formule de Hazen $FND = \frac{n_i - 0,5}{N}$;
- Colonne. 5 : Calcul des logarithmes népériens des pluies classées (on aurait pu aussi bien calculer les logarithmes base 10) ;
- Colonne. 6 : Variables réduites théoriques u_i expérimentales, c'est-à-dire, celles qu'on a tiré de la table de Gauss et qui correspondent aux FND de la colonne 4 ;
- Colonne. 7 : Sachant que les caractéristiques de la série pluviométrique transformée en logarithme sont : $\overline{\ln P} = 6,002$ mm et $S_{\ln P} = 0,30$ mm. $u_i \text{ théo} = \frac{\ln P - \overline{\ln P}}{S_{\ln P}}$;
- Colonne 8 : Pluies théoriques : $P_i \text{ théo} = e^{(\overline{\ln P} + z_i \times S_{\ln P})}$.

Tableau I.14. Ajustement d'une loi log- normale à l'échantillon à la série de pluies annuelles recueillies à une station pluviométrique (Oued FODDA).

1	2	3	4	5	6	7	8
Valeurs Mes.	Valeurs Clas.	Ordre	Fréq. Exp.	ln P	u_i Exp.	u_i Théo.	Valeurs Théo.
456	245	1	0,0208	5,501	-2,037	-1,672	244,94
349	257	2	0,0625	5,549	-1,534	-1,512	256,98
622	297	3	0,1042	5,694	-1,258	-1,028	297,08
520	302	4	0,1458	5,710	-1,054	-0,974	301,87
494	302	5	0,1875	5,710	-0,887	-0,974	301,87
302	315	6	0,2292	5,753	-0,742	-0,831	315,13
448	321	7	0,2708	5,771	-0,610	-0,771	320,86
576	321	8	0,3125	5,771	-0,489	-0,771	320,86
321	342	9	0,3542	5,835	-0,374	-0,557	342,06
449	347	10	0,3958	5,849	-0,264	-2,252	205,82
653	349	11	0,4375	5,855	-0,157	-0,491	348,97
394	394	12	0,4792	5,976	-0,052	-0,087	393,86
587	411	13	0,5208	6,019	0,052	0,057	411,17
681	443	14	0,5625	6,094	0,157	0,307	443,19
257	448	15	0,6042	6,105	0,264	0,344	448,09
347	449	16	0,6458	6,107	0,374	0,350	448,99
315	456	17	0,6875	6,122	0,489	0,400	455,78
245	494	18	0,7292	6,203	0,610	0,671	494,23
342	520	19	0,7708	6,254	0,742	0,841	520,09
297	576	20	0,8125	6,356	0,887	1,181	575,94
302	587	21	0,8542	6,375	1,054	1,245	586,99
321	622	22	0,8958	6,433	1,258	1,438	622,04
411	653	23	0,9375	6,482	1,534	1,602	653,28
443	681	24	0,9792	6,524	2,037	1,742	681,30

Dans la première figure (Fig. I.22), on a porté sur le papier semi-logarithmique (où l'échelle des abscisses est arithmétique et l'échelle des ordonnées logarithmique) en abscisse les variables réduites de la colonne 6 et en ordonnée les pluies correspondantes de la colonne 2.

Dans la deuxième figure (Fig. I.23), on a porté sur le papier de probabilité log normale en abscisse les fréquences expérimentales (FND) de la colonne 4 et en ordonnée les pluies correspondantes de la colonne 2. Ce sont les points expérimentaux.

On peut tracer la droite de Henry soit sur du papier semi-log ou sur du papier de probabilité log normale. La droite de Henry passe par deux points dont les coordonnées sont :

▪ *Sur le papier millimétré*

Abscisses : les variables réduites u_1 et u_2 ;

Ordonnées : les pluies correspondantes P_1 et P_2 .

▪ *Sur le papier de probabilité normale*

Abscisses : les fréquences F_1 et F_2 ;

Ordonnées : les pluies correspondantes P_1 et P_2 .

On prend arbitrairement $F_1 = 0,5$ et $F_2 = 0,9$. On a :

$$F_1 = 0,5 \Rightarrow u_1 = 0 \text{ et } P_1 = e^{(\ln P + z_1 \times s_{\ln P})} = e^{(6+0 \times 0,3)} \Rightarrow P_1 = 403,43 \text{ mm}$$

$$F_2 = 0,9 \Rightarrow u_2 = 1,28 \text{ et } P_2 = e^{(\ln P + z_1 \times s_{\ln P})} = e^{(6+1,28 \times 0,3)} \Rightarrow P_2 = 592,29 \text{ mm}$$

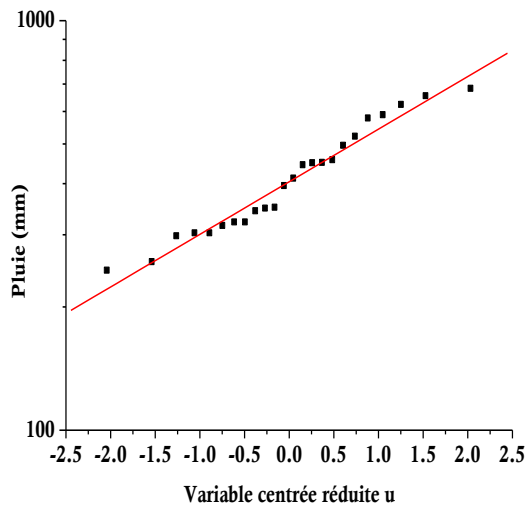


Fig. I.22. Ajustement d'une loi log-normale. Papier semi-logarithmique.

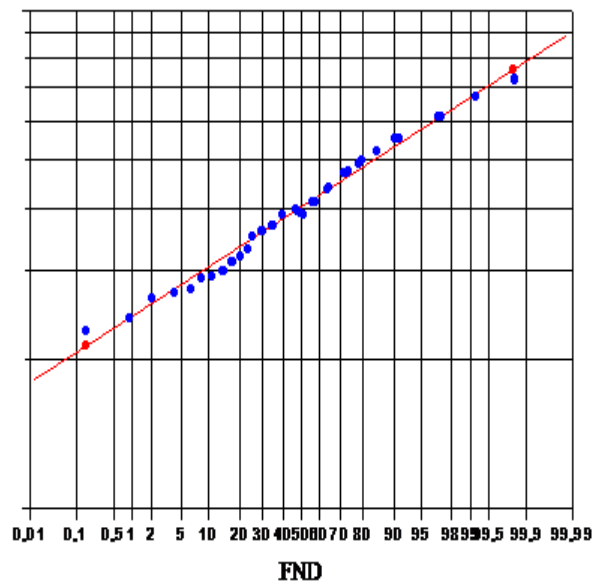


Fig. I.23. Ajustement d'une loi log-normale. Papier de probabilité log normale.

2- Test du χ^2 à 95%.

On établit le tableau I.15 :

- *Colonne. 1* : Numéro de la classe ;
- *Colonne. 2* : Borne inférieure des classes ;
- *Colonne. 3* : Borne supérieure des classes ;
- *Colonnes. 4 et 5* : Logs des bornes inférieures et supérieures des classes ;
- *Colonnes. 6 et 7* : Variables réduites correspondantes à ces bornes ;
- *Colonnes. 8 et 9* : Probabilités au non-dépassement relatives aux bornes respectives, que l'on trouve en utilisant la table de Gauss ;
- *Colonne. 10* : Fréquence expérimentale de chaque classe ;
- *Colonne. 11* : Fréquence théorique de chaque classe $ft_i = N (F(u_2) - F(u_1))$;

- Colonne. 12 : La valeur du $\chi_i^2 = \frac{(fo_i - ft_i)^2}{ft_i}$ pour chaque classe.

Tableau I.15. Application du test du χ^2 .

1	2	3	4	5	6	7	8	9	10	11	12
N°	X ₁	X ₂	ln X ₁	ln X ₂	u ₁	u ₂	F(u ₁)	F(u ₂)	fo _i	ft _i	χ ²
1	− ∞	320	− ∞	5,768	− ∞	-0,780	0	0,218	6	5,22	0,115
2	320	400	5,768	5,991	-0,780	-0,035	0,218	0,486	6	6,44	0,030
3	400	500	5,991	6,215	-0,035	0,710	0,486	0,761	6	6,60	0,055
4	500	+∞	6,215	+∞	0,710	+∞	0,761	1	6	5,73	0,012
									Somme	24	0,213

La somme des $\chi^2 = 0,213$.

On cherche maintenant le $\chi_{v,\alpha}^2$, avec : $\alpha = 0,95$ et $v =$ nombre de degrés de liberté = $k - 1 - r$.

$v=1$ où $k =$ nombre de classes = 4, $r = 2 =$ nombre de paramètres qui déterminent la loi normale.

La table du χ^2 (Annexe III) donné $\chi_{1,0,95}^2 = 3,841$.

Comme le χ^2 calculé est plus petit que celui donné par la table, on accepte donc, l'hypothèse qu'une loi normale ayant une moyenne des logs égale à 6,0 et un écart-type des logs égal à 0,3 représente la population de pluies annuelles (Oued FODDA).

3- Calcul de l'intervalle de confiance de la moyenne à 80%

On sait que le log de la moyenne des pluies est différent de la moyenne des logs des pluies : $\ln \bar{P} \neq \overline{\ln P}$. $\ln \bar{P}$, dans la loi log normale est un quantile comme les autres valeurs de la série des logs. L'intervalle de confiance est donné par la formule suivante :

$$\ln \bar{P} - u_{\frac{1-\alpha}{2}} \frac{S_{\ln P}}{\sqrt{2n}} \sqrt{2 + u_{\frac{1-\alpha}{2}}^2} < x_p < \ln \bar{P} + u_{\frac{1-\alpha}{2}} \frac{S_{\ln P}}{\sqrt{2n}} \sqrt{2 + u_{\frac{1-\alpha}{2}}^2}$$

$$\alpha = 0,8 \Rightarrow 1 - \alpha = 0,2 \Rightarrow \frac{1 - \alpha}{2} = 0,1 \text{ et } u_{\frac{1-\alpha}{2}} = 1,28$$

$$\bar{P} = 422,17 \text{ mm} ; S_p = 128,9 \text{ mm} ; \ln \bar{P} = 6,05 ; S_{\ln P} = 0,3 ; \overline{\ln P} = 6,0$$

$$u_{\ln P} = u_{\ln \bar{P}} = \frac{\ln \bar{P} - \overline{\ln P}}{S_{\ln P}} = \frac{6,05 - 6,0}{0,3} = 0,167$$

$$6,05 - 1,28 \times \frac{0,3}{\sqrt{2 \times 24}} \times \sqrt{2 + (0,167)^2} < \widehat{\ln P} < 6,05 + 1,28 \times \frac{0,3}{\sqrt{2 \times 24}} \times \sqrt{2 + (0,167)^2} \Rightarrow 6,129 < \widehat{\ln P} < 5,971$$

$$\Rightarrow e^{6,129} < \widehat{P} < e^{5,971} \Rightarrow \mathbf{458,9 \text{ mm} < \widehat{P} < 391,9 \text{ mm}}$$

A titre instructif, on calcule l'intervalle de confiance à 80% de la pluie correspondant à la moyenne des logs : $\overline{\ln P} = 6$ ($P = e^6 = 403,3$ mm). La FND de la moyenne des logs est égale à 0,5 d'où $u = 0$ et on a l'intervalle de confiance égal à :

$$6,05 - 1,28 \times \frac{0,3}{\sqrt{2 \times 24}} \times \sqrt{2 + (0)^2} < \widehat{\ln P} < 6,05 + 1,28 \times \frac{0,3}{\sqrt{2 \times 24}} \times \sqrt{2 + (0)^2} \Rightarrow 6,078 < \widehat{\ln P} < 5,922$$

$$\Rightarrow e^{6,078} < \widehat{P} < e^{5,922} \Rightarrow \mathbf{436,3 \text{ mm} < \widehat{P} < 373 \text{ mm}}$$

4- Calcul des pluies décennale et centennale ainsi que leurs intervalles de confiance à 80%.

- *Décennale et de son intervalle de confiance à 80%*

$$\ln P - u_{\frac{1-\alpha}{2}} \frac{S_{\ln P}}{\sqrt{2n}} \sqrt{2 + u_{\ln P}^2} < \hat{\ln P} < \ln P + u_{\frac{1-\alpha}{2}} \frac{S_{\ln P}}{\sqrt{2n}} \sqrt{2 + u_{\ln P}^2}$$

$$\alpha = 0,8 \Rightarrow 1 - \alpha = 0,2 \Rightarrow \frac{1 - \alpha}{2} = 0,1 \text{ et } u_{\frac{1-\alpha}{2}} = 1,28$$

$$\overline{\text{LnP}}_{10} = \overline{\text{LnP}} + u_{\text{LnP}_{10}} \times u_{\text{LnP}};$$

$$S_{\ln P} = 0,3 ; \overline{\text{LnP}} = 6,0$$

$$T = 10 \text{ ans} \Rightarrow \text{FD} = \frac{1}{10} \Rightarrow \text{FD} = 0,10 \Rightarrow \text{FND} = 0,90 \Rightarrow u_{25} = 1,28$$

$$\text{D'où ; } \overline{\text{LnP}}_{10} = \overline{\text{LnP}} + u_{\text{LnP}_{10}} \times u_{\text{LnP}} = 6 + 1,28 \times 0,3 = 6,39 \text{ et } P_{10} = 593,13 \text{ mm}$$

$$6,39 - 1,28 \times \frac{0,3}{\sqrt{2 \times 24}} \times \sqrt{2 + (1,28)^2} < \hat{\ln P} < 6,39 + 1,28 \times \frac{0,3}{\sqrt{2 \times 24}} \times \sqrt{2 + (1,28)^2} \Rightarrow 6,491 < \hat{\ln P} < 6,280$$

$$\Rightarrow e^{6,491} < \hat{\text{LnP}} < e^{6,280} \Rightarrow \mathbf{659,3 \text{ mm} < \hat{P} < 533,6 \text{ mm}}$$

- *Centennale et de son intervalle de confiance à 80%*

$$\ln P - u_{\frac{1-\alpha}{2}} \frac{S_{\ln P}}{\sqrt{2n}} \sqrt{2 + u_{\ln P}^2} < \hat{\ln P} < \ln P + u_{\frac{1-\alpha}{2}} \frac{S_{\ln P}}{\sqrt{2n}} \sqrt{2 + u_{\ln P}^2}$$

$$\alpha = 0,8 \Rightarrow 1 - \alpha = 0,2 \Rightarrow \frac{1 - \alpha}{2} = 0,1 \text{ et } u_{\frac{1-\alpha}{2}} = 1,28$$

$$\overline{\text{LnP}}_{100} = \overline{\text{LnP}} + u_{\text{LnP}_{100}} \times u_{\text{LnP}};$$

$$S_{\ln P} = 0,3 ; \overline{\text{LnP}} = 6,0$$

$$T = 100 \text{ ans} \Rightarrow \text{FD} = \frac{1}{100} \Rightarrow \text{FD} = 0,01 \Rightarrow \text{FND} = 0,990 \Rightarrow u_{25} = 2,33$$

$$\text{D'où ; } \overline{\text{LnP}}_{100} = \overline{\text{LnP}} + u_{\text{LnP}_{100}} \times u_{\text{LnP}} = 6 + 2,33 \times 0,3 = 6,70 \text{ et } P_{100} = 812,35 \text{ mm}$$

$$6,70 - 1,28 \times \frac{0,3}{\sqrt{2 \times 24}} \times \sqrt{2 + (2,33)^2} < \hat{\ln P} < 6,70 + 1,28 \times \frac{0,3}{\sqrt{2 \times 24}} \times \sqrt{2 + (2,33)^2} \Rightarrow 6,806 < \hat{\ln P} < 6,594$$

$$\Rightarrow e^{6,806} < \hat{\text{LnP}} < e^{6,594} \Rightarrow \mathbf{902,9 \text{ mm} < \hat{P} < 730,8 \text{ mm}}$$

Exercice 7.

Par ordre chronologique, les hauteurs de pluies de différentes durées sont représentées dans le tableau ci-dessous (tableau.7).

Tableau.

15 mn	30 mn	60 mn	120 mn	15 mn	30 mn	60 mn	120 mn
8,05	9,72	11,43	17,03	7	11,67	12	17,92
9,4	10,37	10,95	34,21	9,1	8,5	12,67	16,04
8,84	8,73	17,56	13,53	7,33	10	13,75	13,35
7,1	9,19	13,13	12,9	8,1	10,63	11,92	12,3
6,7	26,25	12,13	18,83	7,5	9,34	11,09	17,33
7,83	11,49	11,94	16	7,5	12,38	10,94	13,51
14,23	8,69	32,2	13,1	8,6	10,87	11,3	12,74
6,87	11,7	14,31	12	8,9	8,8	10,63	16,09
7,7	10,27	13,06	19	7,7	10,4	12,25	13
7,32	8,9	18,3	16,63	6,7	11,77	11,3	13,05
9,2	9,7	12,9	14,82	6,9	10,6	12,04	15,7
7	10,23	12,03	13	6,88	9,75	15,1	13,93
7,07	16,26	11,74	12,2				

1. Ajuster les différents échantillons à une loi de Gumbel ;
2. Déterminer les pluies maximales centennale, cinquantenaire, décennale et biennale ;
3. Tracer les courbes Intensités-Durées-Fréquences.

Sari, 2002b

Solution.

1. Tracé des courbes Intensité – Durée – Fréquence

Ajustement des différents échantillons à une loi de Gumbel

Pour cela on calcule les paramètres $\frac{1}{\alpha}$ et x_0 pour chaque série.

$$\frac{1}{\alpha} = 0,78 \times s_p; P_0 = \bar{x} - 0,577\alpha; T = \frac{1}{FD} \Rightarrow FD = \frac{1}{T} \text{ FND} = 1 - \text{FD}$$

$$y = -\ln(-\ln \text{FND}); P_{\max} = \left(\frac{1}{\alpha}\right)y + P_0 \text{ et } I_{\max} = \frac{P_{\max} \times 60 \text{ min}}{\Delta t}$$

Δt	15 min	30 min	60 min	120 min
\bar{P}	7,74	11,05	13,47	15,53
s_p	2,12	3,56	4,34	4,45
$\frac{1}{\alpha}$	1,65	2,78	3,39	3,47
P_0	6,79	9,45	11,51	13,53

Ajustement d'une loi de Gumbel

On établit le tableau I.16 pour P_{\max} de 15 min :

- *Colonne. 1* : Valeurs de l'échantillon, les pluies mesurées ;
- *Colonne. 2* : Pluies classées ;
- *Colonne. 3* : Numéro d'ordre n_i des valeurs classées ;
- *Colonne. 4* : Fréquence au non dépassement expérimentales ($\text{FND}_{\text{exp}} = \frac{n_i - 0,5}{N}$);

- Colonne. 5 : Variables réduites de Gumbel $y_i = -\ln(-\ln FND_i)$;
- Colonne. 6 : Pluies maximales théoriques : $P_i = \frac{1}{\alpha} \times y_i + P_0$.

Tableau I.16. Pmax de 15 min.

1	2	3	4	5	6	1	2	3	4	5	6
Val. Mes.	Val. Clas.	Ordre	Fréq. Exp.	y_i	P.max. Théo.	Val. Mes.	Val. Clas.	Ordre	Fréq. Exp.	y_i	P.max. Théo.
8,05	0,7	1	0,020	-1,364	4,53	7	7,7	14	0,540	0,484	7,59
9,4	6,7	2	0,060	-1,034	5,08	9,1	7,7	15	0,580	0,607	7,79
8,84	6,87	3	0,100	-0,834	5,41	7,33	7,83	16	0,620	0,738	8,01
7,1	6,88	4	0,140	-0,676	5,67	8,1	8,05	17	0,660	0,878	8,24
6,7	6,9	5	0,180	-0,539	5,90	7,5	8,1	18	0,700	1,031	8,49
7,83	7	6	0,220	-0,415	6,10	7,5	8,6	19	0,740	1,200	8,77
14,23	7	7	0,260	-0,298	6,30	8,6	8,84	20	0,780	1,392	9,09
6,87	7,07	8	0,300	-0,186	6,48	8,9	8,9	21	0,820	1,617	9,46
7,7	7,1	9	0,340	-0,076	6,66	7,7	9,1	22	0,860	1,892	9,91
7,32	7,32	10	0,380	0,033	6,84	0,7	9,2	23	0,900	2,250	10,51
9,2	7,33	11	0,420	0,142	7,02	6,9	9,4	24	0,940	2,783	11,38
7	7,5	12	0,460	0,253	7,21	6,88	14,23	25	0,980	3,902	13,23
7,07	7,5	13	0,500	0,367	7,39						

Le graphique (Fig. I.24) est tracé sur du papier millimétré avec en abscisses les variables réduites de Gumbel (colonne 5) et en ordonnées les pluies maximales de durée 15 minutes (colonne 2).

La droite théorique est tracée en joignant 2 points dont l'une des coordonnées est choisie arbitrairement et l'autre coordonnée calculée grâce à l'équation de la droite théorique $P_i = \frac{1}{\alpha} \times y_i + P_0$.

Sur la Fig.I.24 ci-contre, la droite théorique est formée en joignant les points dont les abscisses sont les variables réduites de Gumbel (colonne 5) et les ordonnées les valeurs théoriques (colonne 6).

Dans la seconde figure (Fig. I.25), on porte les points expérimentaux et la droite théorique sur du papier de probabilité de Gumbel. La figure n'est pas à l'échelle, elle est fournie à titre indicatif. Le graphique est tracé sur du papier de probabilité de Gumbel avec en abscisses les FND (colonne 4) et en ordonnées les pluies maximales de durée 15 minutes (colonne 2). La droite théorique est tracée en joignant 2 points dont l'une des coordonnées

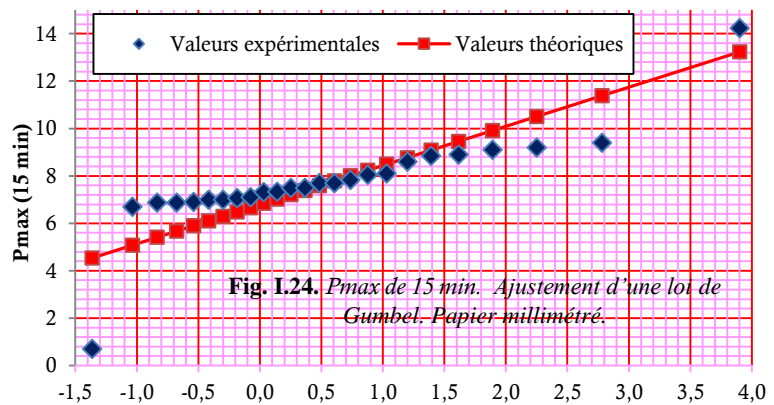


Fig. I.24. Pmax de 15 min. Ajustement d'une loi de Gumbel. Papier millimétré.

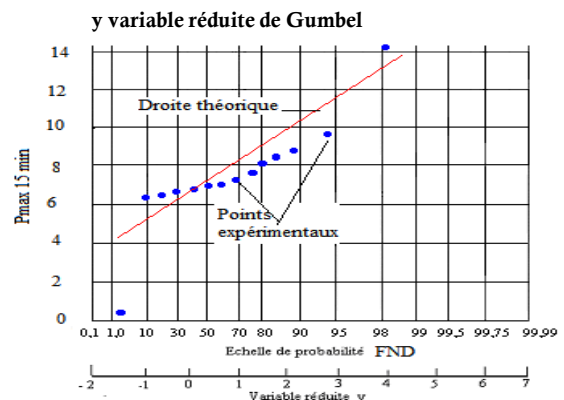


Fig. I.25. Pmax de 15 min. Ajustement d'une loi de Gumbel. Papier de probabilité de Gumbel.

est choisie arbitrairement et l'autre coordonnée calculée grâce à l'équation de la droite théorique $P_i = \frac{1}{\alpha} \times y_i + P_0$ (avec $y_i = -\ln(-\ln(FND_i))$). Sur cette figure, la droite théorique est formée en joignant les points dont les abscisses sont les FND (colonne 4) et les ordonnées les valeurs théoriques (colonne 6).

On établit le tableau I.17 pour Pmax de 60 min :

Tableau I.17. Pmax de 60 min.

1	2	3	4	5	6	1	2	3	4	5	6
Val. Mes.	Val. Clas.	Ordre	Fréq. Exp.	y_i	P.max. Théo	Val. Mes.	Val. Clas.	Ordre	Fréq. Exp.	y_i	P.max. Théo
9,72	8,5	1	0,020	-1,364	5,66	11,67	10,37	14	0,540	0,484	10,79
10,37	8,69	2	0,060	-1,034	6,58	8,5	10,4	15	0,580	0,607	11,13
8,73	8,73	3	0,100	-0,834	7,13	10	10,6	16	0,620	0,738	11,50
9,19	8,8	4	0,140	-0,676	7,57	10,63	10,63	17	0,660	0,878	11,88
26,25	8,9	5	0,180	-0,539	7,95	9,34	10,87	18	0,700	1,031	12,31
11,49	9,19	6	0,220	-0,415	8,30	12,38	11,49	19	0,740	1,200	12,78
8,69	9,34	7	0,260	-0,298	8,62	10,87	11,67	20	0,780	1,392	13,31
11,7	9,7	8	0,300	-0,186	8,93	8,8	11,7	21	0,820	1,617	13,94
10,27	9,72	9	0,340	-0,076	9,24	10,4	11,77	22	0,860	1,892	14,70
8,9	9,75	10	0,380	0,033	9,54	11,77	12,38	23	0,900	2,250	15,69
9,7	10	11	0,420	0,142	9,84	10,6	16,26	24	0,940	2,783	17,17
10,23	10,23	12	0,460	0,253	10,15	9,75	26,25	25	0,980	3,902	20,28
16,26	10,27	13	0,500	0,367	10,46						

Le graphique (Fig. I.26) est tracé sur du papier millimétré avec en abscisses les variables réduites de Gumbel (colonne 5) et en ordonnées les pluies maximales de durée 30 minutes (colonne 2). La droite théorique est tracée en joignant 2 points dont l'une des coordonnées est choisie arbitrairement et l'autre coordonnée calculée grâce à l'équation de la droite théorique $P_i = \frac{1}{\alpha} \times y_i + P_0$.

Sur la figure ci-dessous, la droite théorique est formée en joignant les points dont les abscisses sont les variables réduites de Gumbel (colonne 5) et les ordonnées les valeurs théoriques (colonne 6).

Dans la seconde figure (Fig. I.27), on porte les points expérimentaux et la droite théorique sur du papier de probabilité de Gumbel. La figure n'est pas à l'échelle, elle est fournie à titre indicatif. Le graphique est tracé sur du papier de

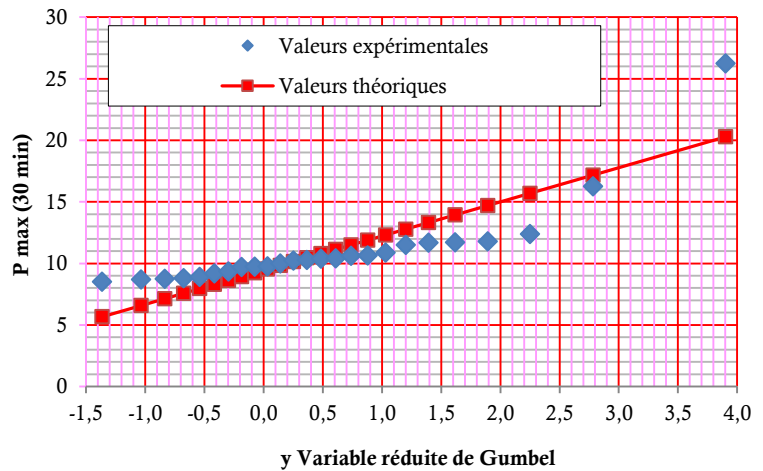


Fig. I.26. Pmax de 60 min. Ajustement d'une loi de Gumbel. Papier millimétré.

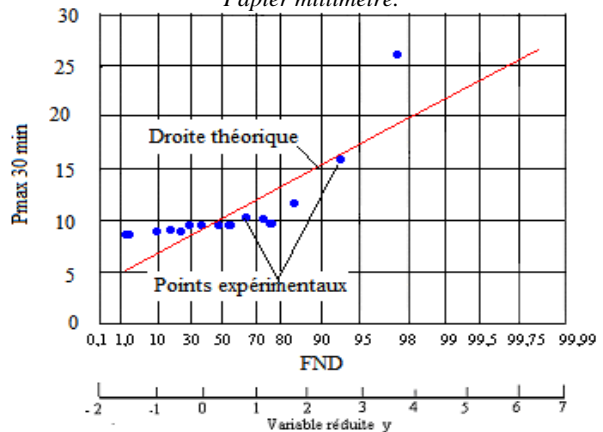


Fig. I.27. Pmax de 60 min. Ajustement d'une loi de Gumbel. Papier de probabilité de Gumbel.

probabilité de Gumbel avec en abscisses les FND (colonne 4) et en ordonnées les pluies maximales de durée 30 minutes (colonne 2). La droite théorique est tracée en joignant 2 points dont l'une des coordonnées est choisie arbitrairement et l'autre coordonnée calculée grâce à l'équation de la droite théorique $P_i = \frac{1}{\alpha} \times y_i + P_0$ (avec ; $y_i = -\ln(-\ln(\text{FND}_i))$). Sur la figure ci-dessus, la droite théorique est formée en joignant les points dont les abscisses sont les FND (colonne 4) et les ordonnées les valeurs théoriques (colonne 6).

On établit le tableau I.18 pour Pmax de 120 min :

Tableau I.18. Pmax de 120 min.

1	2	3	4	5	6	1	2	3	4	5	6
Val. Mes.	Val. Clas.	Ordre	Fréq. Exp.	y_i	P.max. Théo	Val. Mes.	Val. Clas.	Ordre	Fréq. Exp.	y_i	P.max. Théo
17,03	12	1	0,020	-1,364	8,80	17,92	14,82	14	0,540	0,484	15,21
34,21	12,2	2	0,060	-1,034	9,94	16,04	15,7	15	0,580	0,607	15,63
13,53	12,3	3	0,100	-0,834	10,63	13,35	16	16	0,620	0,738	16,09
12,9	12,74	4	0,140	-0,676	11,18	12,3	16,04	17	0,660	0,878	16,57
18,83	12,9	5	0,180	-0,539	11,66	17,33	16,09	18	0,700	1,031	17,10
16	13	6	0,220	-0,415	12,09	13,51	16,63	19	0,740	1,200	17,69
13,1	13	7	0,260	-0,298	12,49	12,74	17,03	20	0,780	1,392	18,36
12	13,05	8	0,300	-0,186	12,88	16,09	17,33	21	0,820	1,617	19,14
19	13,1	9	0,340	-0,076	13,26	13	17,92	22	0,860	1,892	20,09
16,63	13,35	10	0,380	0,033	13,64	13,05	18,83	23	0,900	2,250	21,33
14,82	13,51	11	0,420	0,142	14,02	15,7	19	24	0,940	2,783	23,18
13	13,53	12	0,460	0,253	14,40	13,93	34,21	25	0,980	3,902	27,06
12,2	13,93	13	0,500	0,367	14,80						

Le graphique (Fig. I.28) est tracé sur du papier millimétré avec en abscisses les variables réduites de Gumbel (colonne 5) et en ordonnées les pluies maximales de durée 120 minutes (colonne 2). La droite théorique est tracée en joignant 2 points dont l'une des coordonnées est choisie arbitrairement et l'autre coordonnée calculée grâce à l'équation de la droite théorique $P_i = \frac{1}{\alpha} \times y_i + P_0$.

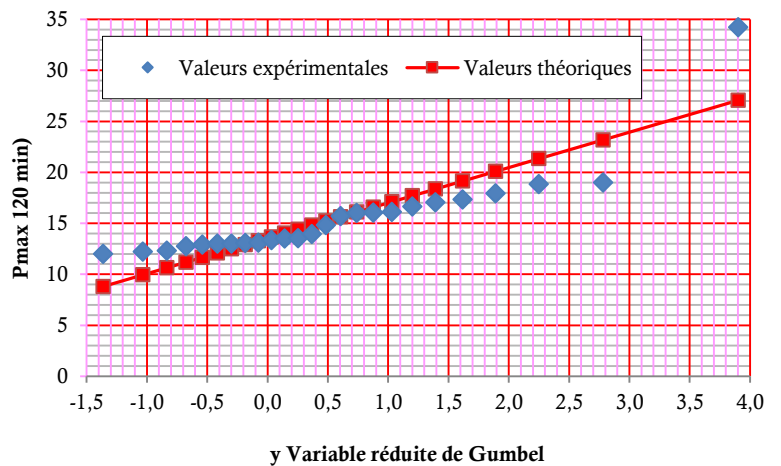


Fig. I.28. Pmax de 120 min. Ajustement d'une loi de Gumbel. Papier millimétré.

Sur la figure ci-contre, la droite théorique est formée en joignant les points dont les abscisses sont les variables réduites de Gumbel (colonne 5) et les ordonnées les valeurs théoriques (colonne 6).

Dans la seconde figure (Fig.I.29) on a porté les points expérimentaux et la droite théorique sur du papier de probabilité de Gumbel. La figure n'est pas à l'échelle, elle est fournie à titre indicatif. Le graphique est tracé sur du papier de probabilité de Gumbel avec en abscisses les FND (colonne 4) et en ordonnées les pluies maximales de durée 120 minutes (colonne 2). La droite théorique est tracée en joignant 2 points dont l'une des coordonnées est choisie arbitrairement et l'autre coordonnée

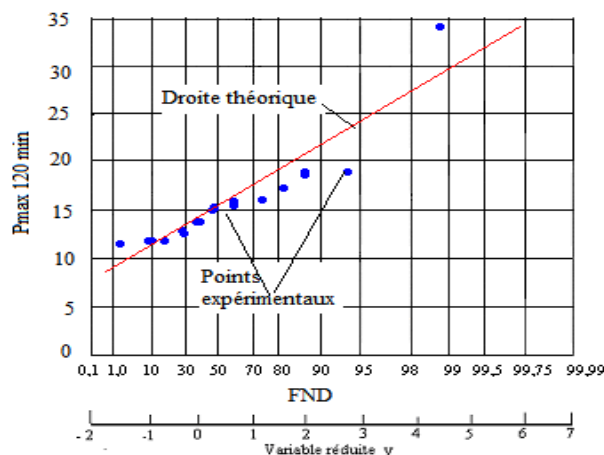


Fig. I.29. Pmax de 120 min. Ajustement d'une loi de Gumbel. Papier de probabilité de Gumbel.

calculée grâce à l'équation de la droite théorique $P_i = \frac{1}{\alpha} \times y_i + P_0$ (avec $y_i = -\ln(-\ln(\text{FND}_i))$). Sur la figure ci-dessus, la droite théorique est formée en joignant les points dont les abscisses sont les FND (colonne 4) et les ordonnées les valeurs théoriques (colonne 6).

2. Courbes intensités – Durées – Fréquences

Dans le tableau ci-dessous on calcule pour chaque durée de pluie les intensités maximales de périodes de retour 2, 10, 50 et 100 ans en utilisant la loi de Gumbel.

Tableau I.19. Pluies maximales et leurs intensités.

Durée (min)	15	30	60	120	15	30	60	120	15	30	60	120	15	30	60	120
T	2	2	2	2	10	10	10	10	50	50	50	50	100	100	100	100
FD	0,5	0,5	0,5	0,5	0,1	0,1	0,1	0,1	0,02	0,02	0,02	0,02	0,01	0,01	0,01	0,01
FND	0,5	0,5	0,5	0,5	0,9	0,9	0,9	0,9	0,98	0,98	0,98	0,98	0,99	0,99	0,99	0,99
Y	0,37	0,37	0,37	0,37	2,25	2,25	2,25	2,25	3,90	3,90	3,90	3,90	4,60	4,60	4,60	4,60
Pmax	7,393	10,5	12,8	14,8	10,5	15,7	19,1	21,3	13,2	20,3	24,7	27,1	14,4	22,2	27,1	29,5
Imax	29,6	20,9	12,8	7,4	42,0	31,4	19,1	10,7	52,9	40,6	24,7	13,5	57,6	44,4	27,1	14,7

En suite sur du papier millimétré (Fig. I.30), on porte pour chaque durée de pluie, les points correspondants aux différentes périodes de retour. Les points de même période de retour sont alors joints les uns aux autres par une même courbe. On obtient ainsi les Courbes Intensités – Durées – Fréquences.

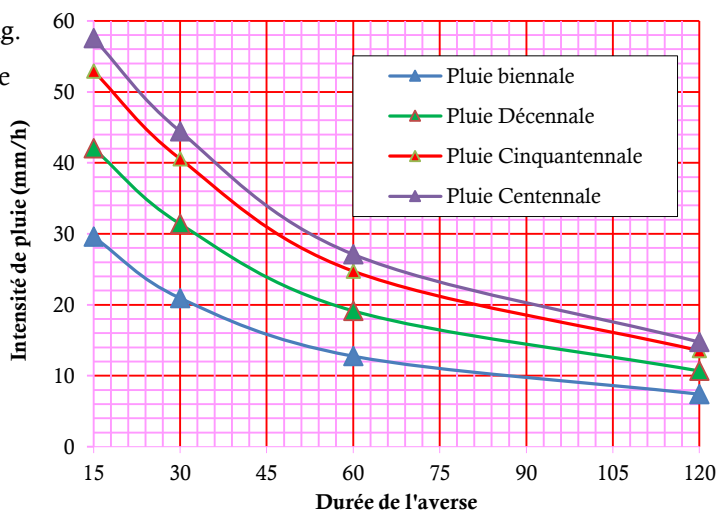
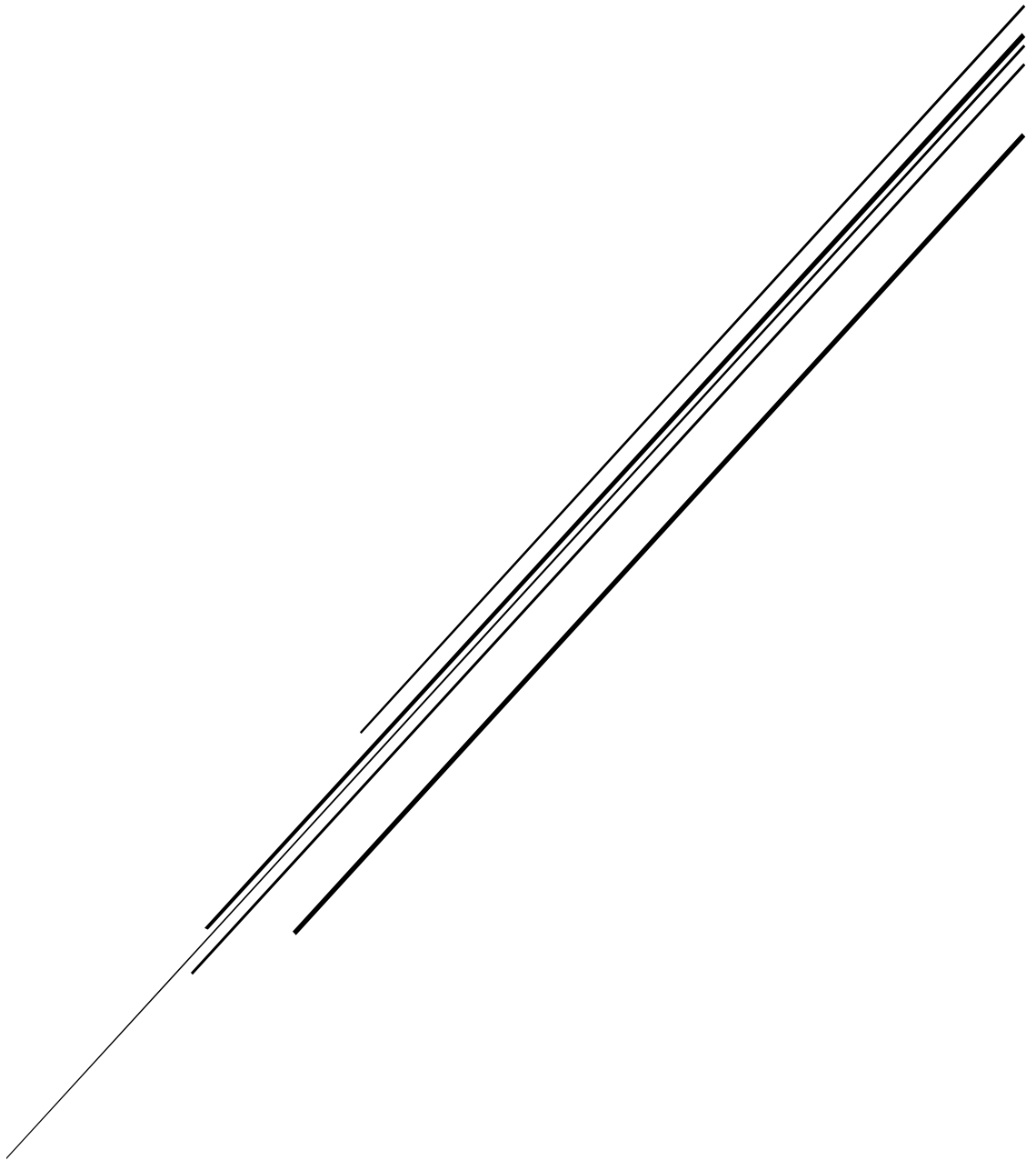


Fig. I.30. Courbes Intensités – Durées – Fréquences.

CORRELATIONS ET ANALYSE DES DONNEES

Chapitre II



CORRELATIONS ET ANALYSE DES DONNEES

II.1. Définitions

En hydrologie, il arrive rarement qu'on étudie les phénomènes comme des variables aléatoires isolées, c'est-à-dire sans prise en compte de leur dépendance vis à vis d'autres phénomènes ou facteurs. Les grandeurs hydrologiques et autres sont mesurées pour déterminer leurs rapports et leur dépendance mutuelle. L'hydrologie et plus précisément l'hydrologie statistique a développé des modèles mathématiques qui décrivent les liaisons (si elles existent !) entre ces phénomènes stochastiques (variables aléatoires). Ces modèles sont appelés modèles de régression.

Dans ce deuxième chapitre, nous allons rappeler dans un premier temps les notions générales de régression et de corrélation. Une présentation des méthodes relatives à la régression sera exposée et en finira par les définitions des notions de séries, variables et processus.

II.1.1. Notions de régression et de corrélation

II.1.1.1. Généralités

Soient deux variables aléatoires X et Y , si on porte sur un graphique les n points de coordonnées x_i, y_i . (i de 1 à n), on obtient quelque chose qui ressemble plus à un nuage de points (Fig. II.1), et ceci pour diverses raisons (Belarbi et al., 2012):

- La liaison n'est pas toujours fonctionnelle (par exemple, cas des lames d'eau précipitées et écoulées annuelles), on dit qu'il existe entre elles une dépendance stochastique et font l'objet d'étude statistique ;
- Les données sont entachées d'erreurs, etc.

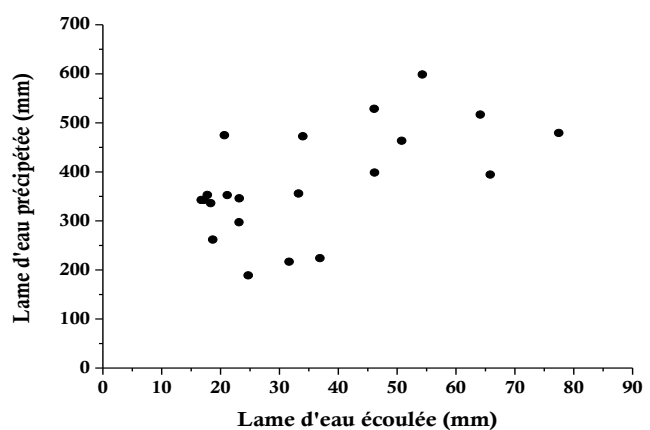


Fig. II.1. Variation annuelle des lames précipitées et écoulées. Bassin versant d'Oued Sebdoou. Période d'étude (1980- 2001).

On peut alors chercher à :

- Schématiser analytiquement cette liaison (par exemple, pour pouvoir facilement estimer une valeur de Y à partir d'une valeur de X) ;
- Caractériser la dépendance entre X et Y par une valeur numérique.

Les applications sont nombreuses et très importantes. Elles peuvent se classer sous quatre catégories :

- Extension dans le temps de séries d'observations hydrologiques qui sont de trop courte durée ou comportent des lacunes ;
- Prévisions de données hydrologiques (apports mensuels, crues, étiages, etc.) en fonction des conditions hydrométéorologiques observées au moment de la prévision ;
- Extension géographique à des bassins non observés des caractéristiques hydrologiques déterminées sur divers bassins versants de régime analogue ;
- Etude de la dépendance entre les valeurs successives d'une série de données hydrologiques (série temporelle).

Dans le cas où il y a une *liaison étroite* entre deux variables X et Y, on dit qu'il y a **corrélation**. Le lieu des points définis par les coordonnées $(x_i; y_i)$ forme la **courbe de régression** de Y en X, où Y est la variable dépendante ou à étudier et X est la variable indépendante ou explicative.

Lorsque la courbe de régression est une droite (dépendance linéaire), les deux droites de régression de Y en X ou de X en Y ne sont pas les mêmes (Fig. II.2)

Régression : C'est une méthode de recherche d'une relation exprimant le lien entre une variable dépendante Y et une ou plusieurs variables dites indépendantes. Dans ce cas la variable dépendante n'est pas forcément une variable aléatoire.

Corrélation : C'est une méthode de recherche de la liaison qui existe entre deux variables aléatoires.

II.1.1.2. Diagrammes de dispersion et choix du modèle de régression

Les observations relatives à deux variables et à plusieurs individus se présentent sous la forme d'une série statistique double c'est-à-dire n couples de valeurs observées (x_i, y_i) rangés éventuellement dans l'ordre croissant de l'une des deux variables (Touaibia, 2004).

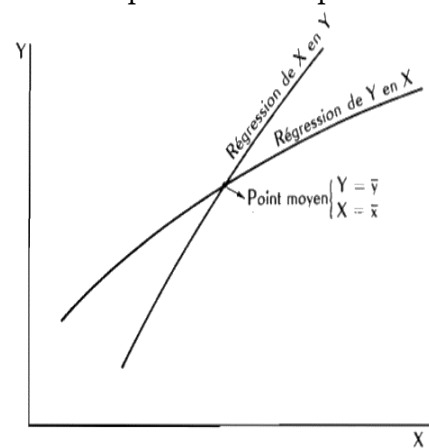


Fig. II.2. Courbes de régression.

Pour étudier simultanément deux séries d'observations, on fait intervenir :

- La notion des moments et de la covariance ;
- Les droites de régression au sens des moindres carrés et les variances résiduelles ;
- Les coefficients de corrélation et de détermination.

La notion de corrélation concerne *l'intensité de la relation* existante entre les deux séries d'observations tandis que la droite de régression donne *l'allure* supposée linéaire de cette relation.

II.1.1.3. Choix du modèle

Le diagramme de régression de Y en fonction de X est formé de points moyens conditionnels (\bar{x}_i, \bar{y}_i) . Il donne une idée de la façon dont varie en moyenne la variable Y dite dépendante en fonction de la variable X dite indépendante.

Lorsque le diagramme de dispersion est linéaire ou approximativement linéaire, on peut s'efforcer de rechercher l'équation de la droite qui s'y ajuste le mieux. Cette droite de régression de Y en X est généralement déterminé par la méthode des *moindres carrés*.

Dans la pratique, on s'efforce toujours de trouver une régression linéaire même s'il faut faire une transformation dans la relation fonctionnelle. Les différentes méthodes existantes sont :

- Le modèle linéaire représenté par l'équation de la droite : $y = a + bx$
- Les modèles curvilinéaires, à savoir :
 1. Le modèle puissance : $y = ax^b$
 2. Le modèle exponentiel : $y = ae^{bx}$
 3. Le modèle parabolique : $y = a + bx + cx^2$

Les coefficients a, b et c sont des constantes à déterminer par la méthode des moindres carrés.

II.1.2. Méthodes relatives à la régression : Cas de la régression à deux variables

II.1.2.1. Méthodes des moindres carrés

Pour la régression linéaire, la droite de régression de Y en X est généralement déterminée par la méthode des moindres carrés, qui consiste à minimiser la somme des carrés des écarts entre les points observés et les points correspondants sur la droite.

Si l'équation de la droite est : $y = a + bx$ et si on dispose d'une série de valeurs (x_i, y_i) , la somme des carrés des écarts à minimiser est :

$$\sum_{i=1}^n E^2 = \sum_{i=1}^n (y_i - y(x_i))^2 = \sum_{i=1}^n (y_i - a - bx_i)^2 = \text{minimum} \quad (\text{II-1})$$

Les valeurs de x_i et y_i sont connues.

Si on pose $y(x_i) = \hat{y}$ avec $\hat{y} = a + bx_i$ on a :

$$\begin{aligned} \sum_{i=1}^n E^2 &= \sum_{i=1}^n (y_i - \hat{y})^2 \\ &= \sum_{i=1}^n (y_i^2 - 2y_i\hat{y} + \hat{y}^2) \\ &= \sum_{i=1}^n (y_i^2 - 2y_i(a + bx_i) + (a + bx_i)^2) \\ &= \sum_{i=1}^n y_i^2 - 2 \sum_{i=1}^n y_i(a + bx_i) + \sum_{i=1}^n (a + bx_i)^2 \end{aligned} \quad (\text{II-2})$$

En développant cette expression, en recherchant les dérivées partielles par rapport à a et b , et obtient la pente b et l'ordonnée à l'origine a , soit :

$$b = \frac{\sum_{i=1}^n (x_i y_i) - \bar{y} \sum_{i=1}^n x_i}{\sum_{i=1}^n x_i^2 - \bar{x} \sum_{i=1}^n x_i} = \frac{\sum_{i=1}^n (x_i y_i) - n \bar{x} \bar{y}}{\sum_{i=1}^n x_i^2 - n \bar{x}^2} \quad (\text{II-3})$$

Et,

$$a = \bar{y} - b \bar{x} \quad (\text{II-4})$$

La droite de régression passe par le point (\bar{x}, \bar{y}) .

La pente b peut être exprimée d'une autre façon :

$$b = \frac{\sum_{i=1}^n (x_i y_i) - \frac{1}{n} (\sum_{i=1}^n x_i) (\sum_{i=1}^n y_i)}{\sum_{i=1}^n x_i^2 - \frac{1}{n} (\sum_{i=1}^n x_i)^2} = \frac{\text{cov}(x,y)}{S_x^2} \quad (\text{II-5})$$

L'équation de la droite de régression de y en x , s'écrira :

$$y = \frac{\text{cov}(x,y)}{S_x^2} (x - \bar{x}) + \bar{y} = b_{y/x} (x - \bar{x}) + \bar{y} \quad (\text{II-6})$$

Avec ; $b_{y/x}$: coefficient de régression de y en x

A. Notion des moments et de la covariance

La notion généralisée des moments et exprimée par la formule II-7 :

$$a_{k,l} = \frac{1}{n} \sum_{i=1}^n [(x_i - c)^k (y_i - d)^l] \quad (\text{II-7})$$

$a_{k,l}$ est le nombre non centré d'ordre k en x par rapport à c et d'ordre l en y par rapport à d

En posant : $c = \bar{x}$ et $d = \bar{y}$, on obtient les moments centrés $m_{k,l}$ appelés aussi *moment par rapport aux moyennes*.

La notion généralisée des moments centrés permet de définir la covariance $m_{1,1}$ qui caractérise simultanément les deux séries d'observations.

$$\text{Cov}(x, y) = m_{11} = \frac{1}{n} \sum_{i=1}^n [(x_i - \bar{x})(y_i - \bar{y})] \quad (\text{II-8})$$

B. Régression curvilinéaire

Dans de nombreux cas, il existe une relation entre les points observés. Cette relation peut être linéaire et on peut procéder à l'ajustement d'une courbe de régression au nuage de points observés. Il existe des tests de linéarité permettant de contrôler d'une manière rigoureuse si l'ajustement est satisfaisant ou non. Lorsque la courbe est ajustée, deux problèmes se posent :

- Le choix de l'équation de régression de cette courbe
- Les paramètres intervenant dans cette courbe

Les paramètres d'ajustement d'un polynôme peuvent être déterminés par la méthode des moindres carrés :

$$Y = b_0 + b_1X + b_2X^2 + \dots + b_kX^k \quad (\text{II-9})$$

Il faut minimiser la quantité :

$$\sum_{i=1}^n (b_0 + b_1X + b_2X^2 + \dots + b_kX^k)^2 \quad (\text{II-10})$$

On obtient $(k + 1)$ équations normales.

D'une façon générale, les équations dites normales s'écrivent :

$$\left\{ \begin{array}{l} b_0 n + b_1 \sum_{i=1}^n x_i + b_2 \sum_{i=1}^n x_i^2 + \dots + b_k \sum_{i=1}^n x_i^k = \sum_{i=1}^n y_i \\ b_0 \sum_{i=1}^n x_i + b_1 \sum_{i=1}^n x_i^2 + b_2 \sum_{i=1}^n x_i^3 + \dots + b_k \sum_{i=1}^n x_i^{k+1} = \sum_{i=1}^n x_i y_i \\ \dots \dots \dots \\ b_0 \sum_{i=1}^n x_i^k + b_1 \sum_{i=1}^n x_i^{k+1} + b_2 \sum_{i=1}^n x_i^{k+2} + \dots + b_k \sum_{i=1}^n x_i^{2k} = \sum_{i=1}^n x_i^k y_i \end{array} \right. \quad (\text{II-11})$$

Si $k=1$, on a une régression linéaire ;

Si $k=2$, on a une régression quadratique.

Le rapport de corrélation ou rapport de corrélation non linéaire ou aussi coefficient de corrélation non linéaire $\eta_{y,x}$, joue dans le cas de la régression curvilinéaire un rôle analogue à celui du coefficient de corrélation r dans le cas d'une régression linéaire.

$$\eta_{y,x} = \frac{\sum_{i=1}^n n_i (\bar{y}_i - \bar{y})}{n S_y^2} \quad (\text{II-12})$$

Ce paramètre est toujours compris entre 0 et 1.

$$\eta_{y,x} \geq r^2$$

La différence $\eta_{y,x}$ et r^2 est une mesure de la non linéarité de la régression de y en x .

Cas du modèle parabolique : $Y = a + bX + cX^2$

Elle ne peut être traitée qu'après avoir résolu un système de trois équations dites équations normales.

$$\begin{cases} \sum_{i=1}^n y_i = an + b \sum_{i=1}^n x_i + c \sum_{i=1}^n x_i^2 \\ \sum_{i=1}^n x_i y_i = a \sum_{i=1}^n x_i + b \sum_{i=1}^n x_i^2 + c \sum_{i=1}^n x_i^3 \\ \sum_{i=1}^n x_i^2 y_i = a \sum_{i=1}^n x_i^2 + b \sum_{i=1}^n x_i^3 + c \sum_{i=1}^n x_i^4 \end{cases} \quad (\text{II-13})$$

Dans cette manière, on peut résoudre le système et calculer les coefficients a, b et c.

Cas du modèle puissance $Y = aX^b$ (II-14)

Par une transformation logarithmique, on a : $\log Y = \log a + b \log X$ (II-15)

Cas du modèle exponentiel $Y = ae^{bX}$ (II-16)

Par une transformation logarithmique, on a : $\log Y = \log a + \log e^{bX} = \log a + bX$ (II-17)

Dans les deux cas, les coefficients a et b sont déterminés par la méthode des moindres carrés.

II.1.3. Méthodes relatives à la corrélation

II.1.3.1. Coefficient de corrélation linéaire Coefficient de Bravais Pearson r

Ce coefficient est donné par :

$$r = \frac{\text{cov}(x,y)}{S_x S_y} = \frac{m_{11}}{S_x S_y} \quad (\text{II-18})$$

Avec ; S_x et $S_y \neq 0$

$$\text{Et, } \text{cov}(x,y) = \sum_{i=1}^n [(x_i - \bar{x})(y - \bar{y})] \quad (\text{II-19})$$

Le coefficient de corrélation est compris entre $-1 < r < +1$

Si

$r = \pm 1$, tous les points observés se trouvent sur la droite, l'ajustement est parfait ;

$r = 0$ ou $r \Rightarrow 0$, la dispersion des points est très grande, l'ajustement n'est pas à considérer ;

$-1 < r < +1$, le nuage des points est plus ou moins allongé, il convient de chercher la valeur de r.

L'interprétation de la valeur trouvée de r dépend du risque d'erreur admise et du nombre de degré de liberté qui reste lié à la taille de l'échantillon. La table de Fisher permet de donner le seuil critique admissible pour une erreur α donnée.

Le coefficient de corrélation est indépendant de toute transformation linéaire positive.

Dans le cas où on applique la notion de régression de Yen X, le coefficient de corrélation est directement lié au rapport de la variance résiduelle à la variance marginale. Par définition, la variance marginale est :

$$S_{y/x}^2 = S_y^2(1 - r^2) \quad (\text{II-20})$$

$$\text{Et ; } R^2 = 1 - \frac{S_{y/x}^2}{S_y^2} \quad (\text{II-21})$$

Cette quantité R^2 s'appelle coefficient de détermination.

II.1.3.2. Coefficient de détermination R^2

Il représente le carré du coefficient de corrélation. Il mesure le rapport de la variation expliquée par la régression à la variation totale. Il indique le pourcentage de la variation totale de la variable Y autour de sa moyenne qui est expliquée par la régression en X.

Si ;

$R^2 \approx 1$ cela veut dire que toute la variable est expliquée par la régression et on a une dépendance fonctionnelle.

$R^2 \approx 0$ la régression n'explique rien, les variables sont indépendantes et il n'existe aucune dépendance fonctionnelle.

II.1.4. Complément statistique des données hydrologiques

II.1.4.1. Estimation des caractéristiques empiriques d'une série étendue

Soient deux séries hydrologiques de hauteurs annuelles de précipitations

X : Série longue constituant la série de base à n observations

Y : Série courte ou série à étudier à k observations ($k < n$)

Les estimations des paramètres statistiques et des valeurs annuelles d'une série X de n valeurs observées peuvent être estimées sans pour autant passer par la méthode classique.

A. Moyenne estimée par l'extension

$$\bar{\hat{y}} = \bar{y}_k + r_{xy} \frac{S_y}{S_x} (\bar{x}_n - \bar{x}_k) \quad (\text{II-22})$$

$\bar{\hat{y}}$: Estimation de la moyenne de la série des y (dont l'espérance mathématique est \bar{y} (toujours inconnue) à partir de la 1^{er} estimation de \bar{y}_k , des autres paramètres statistiques des échantillons des valeurs observées et du coefficient de corrélation entre x et y.

On note que l'équation (II-23) est de même forme que l'équation (II-22) que l'on utilise si on veut avoir une meilleure estimation de la moyenne des pluies y. L'équation (II-22) est utilisée

seulement pour reconstituer les $(n - k)$ années observées à l'aide des valeurs correspondantes de x .

B. Variance estimée par l'extension

$$\hat{\sigma}^2 = S_y^2 + r_{xy}^2 \frac{S_y}{S_x} (S_x^2 - S_x^2) \quad (\text{II-23})$$

$\hat{\sigma}^2$: Estimation de la variance de y (dont l'espérance mathématique est σ_y^2) à partir de la 1ère estimation de S_y^2 et des estimations de la variance de x et du coefficient de corrélation x et y ;

S_x^2 : Variance de x à partir de l'échantillon de n valeurs.

Le résultat trouvé par la formule (II-23) est celui que l'on retrouverait en appliquant la formule de définition de la variance :

$$S_y^2 = \frac{\sum_{i=1}^n (y_i - \bar{y}_n)^2}{n-1} \quad (\text{II-24})$$

Et qu'en appliquant à la série de k valeurs y_i observées et des $(n - k)$ valeurs estimées par la formule (II-41).

L'estimation de la variance σ_y^2 étant surement trop faible, la dispersion des valeurs estimées est plus faible que celle des valeurs observées.

Si σ_y^2 est plus faible que S_y^2 ; on conserve la valeur de S_y^2 à la place de σ_y^2 sauf si le coefficient de corrélation r_{xy} est supérieur ou égal à 0,95 car $\hat{\sigma}_y^2$ donne une meilleure estimation de S_y^2 puisqu'elle a la valeur que donnerait un échantillon de n valeurs observées.

C. Estimation du coefficient de corrélation

$$\hat{r} = r_{xy} \frac{S_y \hat{S}_x}{S_x \hat{\sigma}_y} \quad (\text{II-25})$$

\hat{r} : Estimation du coefficient de corrélation entre x et y (dont l'espérance mathématique est r à partir de la 1ère 1ère estimation r_{xy} ; des écarts types des échantillons de valeurs observées et de l'estimation σ_y^2 définie ci-dessous.

Remarque : En pratique, comme il est difficile de tester la signification relative de \hat{r} et de r_{xy} , on conserve le coefficient de corrélation expérimental \hat{r}_{xy} .

D. Valeurs estimées par extension

$$y_x = \bar{y}_k + r_{xy} \frac{S_y}{S_x} (x - \bar{x}_k) \quad (\text{II-26})$$

Avec ;

y_x : Moyenne conditionnelle de y liée à x ;

\bar{x}_k et \bar{y}_k : Moyennes arithmétiques de x et de y calculées respectivement à partir de K valeurs observées simultanément ;

r_{xy} : Coefficient de corrélation entre x et y sur k années d'observations simultanées ;

S_x et S_y : écart type de x et de y sur k années observées simultanément.

II.2. Séries, variables et processus

II.2.1. Définitions

II.2.1.1. Série chronologique

On appelle série chronologique (série temporelle) une suite d'observations numériques d'une grandeur effectuées à intervalles réguliers au cours du temps. De façon plus mathématique, une série chronologique sera mieux interprétée par la définition suivante :

La suite d'observations $(t, y_t \in T)$ d'une variable Y à différentes dates t est appelée série chronologique. Habituellement, T est dénombrable, de sorte que $t = 1, \dots, T$.

Remarque

En mathématiques, la définition de série chronologique ci-dessus correspond à la définition d'une suite $\{u_n\}_{n \in \mathbb{I}}$, tandis qu'on nomme série la suite définie à partir de la somme des termes de la suite : $s_n = \sum_{i=0}^n u_i$.

Une série chronologique est donc toute suite d'observations correspondant à la même variable. En pratique, la série chronologique $\{y_t\}$ est donnée sous forme d'un tableau bidimensionnel où la date peut être remplacée par le numéro d'observation (*Belarbi et al., 2012*).

A. Représentation graphique

La représentation graphique des observations est une étape indispensable avant d'entreprendre une analyse plus technique de la chronique. Les points (t, y_t) $t \in T$, avec $T = \{t_1, t_2, \dots, t_n\}$, sont représentés dans un système d'axes orthogonaux (échelles arithmétiques). Ils sont joints chronologiquement par des segments de droite pour faciliter la visualisation.

Cette représentation permet d'apprécier l'évolution du phénomène, par exemple lorsqu'une série est stable autour de sa moyenne, on parle de *série stationnaire*. Inversement, on trouve aussi des *séries non stationnaires*. Lorsqu'une série croît sur l'ensemble de l'échantillon et donc possède une moyenne qui n'est pas constante, on parle de *tendance*. Enfin, lorsqu'on observe des phénomènes qui se reproduisent à des périodes régulières, on parle de *phénomène saisonnier ou cyclique*.

B. Concepts mathématiques pour aborder les séries chronologiques

Pour que la définition d'une série temporelle soit précise, il est nécessaire de faire référence à la notion de *processus stochastique* (ou *aléatoire*). Cependant, avant d'aborder le concept mathématique d'un processus stochastique, nous commençons par évoquer celle d'une *variable aléatoire*.

II.2.1.2. Variables aléatoires

Soit (Ω, M, Pr) un espace de probabilité, où Ω est l'espace des évènements, M est une tribu adaptée à Ω (c'est l'ensemble qui contient les combinaisons possibles d'évènements) et Pr est une mesure de probabilité définie sur M .

Une variable aléatoire réelle (V.A.R) est une fonction à valeurs réelles $y: \Omega \rightarrow \mathbb{R}$ telle que pour tout réel c . $A_c = \{\omega \in \Omega / y(\omega) \leq c\} \in M$

En d'autres termes A_c est un évènement dont la probabilité est définie en termes de Pr . La fonction $F: \mathbb{R} \rightarrow [0,1]$ définie par $F(c) = Pr(A_c)$ est *la fonction de distribution de y* .

II.2.1.3. Processus stochastiques

Soit T un ensemble d'indexation dénombrable contenu dans l'ensemble des entiers naturels ou dans celui des entiers relatifs.

Un processus stochastique est une fonction à valeurs réelles. $y: T \times \Omega \rightarrow \mathbb{R}$

Telle que pour tout $t \in T$ donné $y_t(\cdot)$ soit une variable aléatoire.

En d'autres termes, un *processus stochastique* est une suite ordonnée de variables aléatoires $\{y_t(\omega), \omega \in \Omega, t \in T\}$ telle que pour tout $t \in T$, y_t soit une variable aléatoire sur Ω et que pour tout $\omega \in \Omega$, $y_t(\omega)$ soit une réalisation de processus stochastique sur l'ensemble d'indexation T .

Une série chronologique $\{y_t\}_{t \in T}$ est une réalisation d'un processus stochastique $\{y_t\}$.

A. Quelques exemples de processus

i. Processus stationnaire au sens strict

Un processus est dit stationnaire au sens strict ou stationnaire au premier ordre, si pour toutes valeurs h_1, h_2, \dots, h_n la distribution jointe de la suite $\{y_t, y_{t+h_1}, \dots, y_{t+h_n}\}$ dépend uniquement des intervalles de temps h_1, h_2, \dots, h_n et non de t .

La stationnarité stricte implique que tous les moments soient indépendants du temps.

ii. *Processus stationnaire au sens faible*

Un processus est dit stationnaire au sens faible ou stationnaire au second ordre, si les premier (moyenne ou espérance mathématique) et second (variance et auto-covariances) moments du processus existent et sont indépendantes de t.

La stationnarité est une propriété de stabilité. La distribution de y_t est identique à celle de y_{t-1} . La série oscille autour de sa moyenne avec une variance constante. Le lien entre y_t et y_{t-h} ne dépend alors que de l'intervalle h et non de la date t .

iii. *Processus bruit blanc*

Un bruit blanc (White Noise) est un processus stationnaire au sens faible de moyenne zéro et qui est dynamiquement non corrélé. $u_t \approx \text{WN}(0, S^2)$.

Ainsi $\{u_t\}$ est un bruit blanc si pour tout $t \in T$: $E[u_t] = 0$; $E[u_t^2] = S^2 < \infty$ avec u_t et u_{t-h} indépendantes si $h \neq 0$; t et $(t - h) \in T$.

iv. *Processus gaussiens*

Ce sont ceux qui ont une loi chronologique normale, les moments des deux premiers ordres suffisent donc à le caractériser.

II.2.2. Valeurs caractéristiques d'une série chronologique

Pour figurer l'évolution dans le temps de certains phénomènes aléatoires, les statisticiens s'accordent pour décomposer une série chronologique $\{y_t\}$ en une somme de deux composantes aux propriétés différentes, soit :

$$y_t = TD_t + S_t \quad t = 1, \dots, n \quad (\text{II-27})$$

Où,

- TD_t : désigne la partie déterministe de la série ;
- S_t : la partie stochastique (en général, de moyenne nulle par hypothèse).

En pratique, de multiples spécifications sont possibles pour la composante déterministe TD_t , celle qui se retrouve la plus couramment utilisée fait de TD_t une fonction polynomiale du temps, linéaire ou non linéaire. Dans les séries chronologiques de données hydro-climatologiques, la spécification la plus utilisée fait de TD_t un polynôme de premier degré en t :

$$TD_t = (a \cdot t) + b \quad t = 1, \dots, n \quad (\text{II-28})$$

Où, $(a, b) \in \mathbb{R}^2$

En associant les variations accidentelles ε_t au modèle précédent on peut alors représenter cette série chronologique sous la forme suivante :

$$y_t = (a \cdot t) + b + \varepsilon_t \quad t = 1, \dots, n \quad (\text{II-29})$$

Dans ce modèle ε_t désigne *les irrégularités* ou *mouvement résiduel* (dite aussi, *résidu* ou *bruit*). Il regroupe ce qui n'a pas été pris en compte par la composante TD_t qui représente la fonction déterministe de la tendance de la série temporelle $\{y_t\}$. Ce sont des variations de faible intensité, de courte durée et de nature aléatoire. En effet, elles ne sont pas clairement apercevables dans les graphiques à cause de leur faible intensité par rapport aux autres composantes. Elles apparaissent clairement seulement après « *l'enlèvement du signal* ». La composante TD_t représente l'évolution à long terme de la suite $\{y_t\}$ étudiée, elle traduit le comportement *moyen* de la série.

La partie stochastique, se décompose comme suit :

$$S_t = TS_t + C_t \quad t = 1, \dots, n \quad (\text{II-30})$$

Où,

- TS_t : Un processus caractérisant la composante stochastique de la fonction de tendance associée à la série temporelle $\{y_t\}$;
- C_t : La composante cyclique.

Le plus souvent TS_t est une fonction du passé de la série chronologique $\{y_t\}$:

$$y_t = F(y_{t-1}, y_{t-2}, y_{t-3}, \dots) \quad t = 1, \dots, n \quad (\text{II-31})$$

Cette dernière formulation suppose que la valeur de la grandeur y_t à la date t dépend des valeurs obtenues par celle-ci aux dates antérieures. Comme la *mémoire* du phénomène ayant généré la série chronologique $\{y_t\}$ se trouve dans le passé le plus récent de la série, il n'est pas utile de remonter trop loin dans le temps. On exprime donc ces constatations par le fait que la composante TS_t représente la partie stochastique de la fonction de tendance associée à cette série temporelle.

Désignons à présent T_t la fonction de tendance (trend) et définissons

$$T_t = TD_t + TS_t \quad (\text{II-32})$$

En d'autres termes, cette fonction est la somme des tendances à la fois déterministe et stochastique. On se trouve donc avec la modélisation suivante d'une série chronologique :

$$y_t = TD_t + TS_t + C_t + \varepsilon_t \quad t = 1, \dots, n \quad (\text{II-33})$$

Cette modélisation n'est en général pas unique puisque des conditions plus générales pourraient être spécifiées. En effet, hormis l'autocorrélation (phénomène de persistance) que l'on peut

trouver entre les données de la série, chacune des composantes de $\{y_t\}$, peut contenir une composante saisonnière. Cette dernière correspond à un phénomène qui se répète à intervalle de temps réguliers (périodique), elle permet simplement de distinguer à l'intérieur d'une même période une répartition stable dans le temps d'effets positifs ou négatifs qui se compensent sur l'ensemble de la période.

En résumé, nous considérons une série chronologique comme issue de la composition de :

TD_t : représente la tendance déterministe ; TS_t : représente la tendance stochastique ;

C_t : représente la variation cyclique ; Sa_t : représente la variation saisonnière ;

R_t : tient compte du phénomène de persistance ; ε_t : tient compte des variations résiduelles purement aléatoires.

Remarque

Il faut juste préciser que dans certains cas, lorsqu'on représente graphiquement une série temporelle, on peut observer une composante déterministe qui se présente sous la forme d'une tendance ou d'un cycle saisonnier. Comme on peut également constater la présence de ces deux composantes en même temps. Parfois, la tendance ou la saisonnalité n'est pas toujours facile à déceler. Il peut même arriver que des séries ne contiennent aucune de ces composantes déterministes. Il se peut aussi que le trend n'apparaisse qu'après avoir transformé les données par une fonction, par exemple, logarithmique

II.2.3. Lissage des séries chronologiques

Pour modéliser l'évolution chronologique d'une série d'observation, on peut par exemple chercher une fonction paramétrique qui ressemble à l'allure générale de la série et estimer les paramètres de cette fonction afin d'ajuster le mieux possible les observations. Les fonctions les plus utilisées sont des fonctions, de type linéaire, polynomiales et dans d'autres cas de type exponentiel. Pour les séries chronologiques de données hydro-climatologiques la spécification la plus utilisée fait de cette évolution chronologique un polynôme de premier degré. La série chronologique s'écrit sous la forme suivante (*Belarbi et al., 2012*):

$$y_t = at + b + \varepsilon_t \quad (\text{II-34})$$

Où a et b sont déterminés par la méthode des moindres carrés.

Il existe d'autres méthodes (non paramétriques) qui permettent de caractériser le mouvement général d'une série chronologique de variable hydro-climatologique. Parmi les méthodes les plus utilisées on peut citer la méthode des moyennes mobiles.

Cette méthode est une technique qui permet de lisser (ou filtrer) directement la série, sans hypothèse à priori sur la forme du modèle sous-jacent. La méthode est donc valable quel que soit le modèle de décomposition. Pour cette raison, on peut classer ce type de lissage dans les méthodes non paramétriques.

La méthode a pour but de réduire l'influence des variations accidentelles et d'éliminer l'effet des fluctuations de très courte période, mais quelquefois elle introduit des variations quasi périodiques de fréquence basse. On constate que l'on affaiblit les composants aléatoires et périodiques mais on ne touche pas au mouvement général de la série.

Cet outil simple à mettre en œuvre a l'inconvénient de créer au sein de la série des corrélations internes appelées *effet de Slutsky-Yule*. L'opérateur moyen mobile appliqué à une série aléatoire peut induire une série plus organisée. Le principe de cette méthode consiste à remplacer la série d'origine par une série de valeurs moyennes mobiles définies par :

$$y_k = \frac{\sum_{i=-n}^n F_i y_{k+i}}{\sum_{i=-n}^n F_i} \quad (\text{II-35})$$

Avec, F_i est une fonction de pondération.

Dans ce cas (moyennes mobiles pondérées) le procédé le plus utilisé est celui de Spencer à quinze ou vingt et un points, cette application nécessite d'énormes opérations particulièrement si la série d'observation est assez longue et présente l'inconvénient d'éliminer les informations au début et à la fin de la série.

Le cas le plus simple est les moyennes mobiles à $(2n + 1)$ points non pondérés, c'est-à-dire $F_i = 1$.

$$y_k = \frac{1}{2n+1} \sum_{i=-n}^n (y_{k+i}) \quad (\text{II-36})$$

Avec cette formule on élimine les variations périodiques de périodicité $(2n + 1)$.

II.3. Test d'homogénéité

II.3.1. Approche générale

L'étude des séries d'observations met parfois en évidence des changements de valeurs qui surviennent brusquement et se maintiennent ultérieurement ou qui se traduisent par une modification progressive mais systématique de ces valeurs, on dit alors que ces séries ne sont pas homogènes.

La constatation de l'hétérogénéité d'une série présente une difficulté variable selon l'élément étudié.

Une série d'observation est considérée comme homogène lorsque les conditions de mesure ou d'autres modifications éventuelles n'auront pas altéré sensiblement les observations. Autrement dit, une série temporelle sera considérée comme homogène si la loi de distribution de l'élément (considéré comme une variable aléatoire) n'a pas changé pendant toute la période étudiée.

Les hétérogénéités que causent les modifications touchant les réseaux d'observation prennent la forme de *discontinuités abruptes*, de *variations graduelles* ou de *changements de variabilité*.

Les *discontinuités abruptes* résultent principalement de (*Belarbi et al., 2012*):

- 1- remplacements d'instruments ;
- 2- changements d'emplacement et d'exposition des instruments ;
- 3- changements d'emplacement des stations ;
- 4- changements touchant les calculs de moyennes et les procédures de réduction des données ainsi que de l'application de nouvelles corrections.

Les hétérogénéités qui prennent la forme d'un *effet croissant graduellement* peuvent découler :

- 1- d'une modification des alentours de la station ;
- 2- de l'urbanisation, ainsi que de variations progressives des caractéristiques instrumentales.

Les *changements de variabilité* sont causés par les défauts de fonctionnement des instruments. Viennent aussi s'ajouter à ces causes d'hétérogénéité, les modifications des horaires d'observation, une fréquence insuffisante des activités régulières d'inspection, de maintenance et d'étalonnage et les procédures d'observation non adaptées. Il importe donc de disposer de renseignements sur toutes les hétérogénéités, leur type et surtout le moment où elles apparaissent.

Il y a lieu de noter que ces listes d'erreur sont presque exhaustives et que toutes ces erreurs n'arrivent pas simultanément. Cependant, les documents élaborés par les observateurs et les hydrologues restent notre référence et sont à la base du développement hydraulique du pays. Les listes ci-dessus sont beaucoup plus pédagogiques.

Les erreurs dans les séries de mesures modifient le caractère aléatoire des phénomènes et les conditions de leurs avènements. Si ces conditions changent cela veut dire que les données mesurées ne proviennent pas de la même population et que la série de mesures n'est pas homogène. Avant de pouvoir étudier statistiquement ces séries, il y a lieu donc, au préalable, de les rendre homogènes.

II.4. Mise en application

Exercice 1.

Soient les séries d'observations des pluies annuelles en trois stations A, B, et C données par le tableau suivant:

Tableau.

Année	Précipitations annuelles (mm)			Année	Précipitations annuelles (mm)		
	Station A	Station B	Station C		Station A	Station B	Station C
1910	222	-	185	1925	266	136	-
1911	150	-	192	1926	327	184	-
1912	281	-	351	1927	310	252	-
1913	210	-	240	1928	360	285	-
1914	172	-	250	1929	413	318	-
1915	260	229	332	1930	258	206	250
1916	252	208	-	1931	225	156	242
1917	321	204	-	1932	308	237	221
1918	382	359	-	1933	-	209	320
1919	364	236	-	1934	-	209	325
1920	193	-	250	1935	349	-	250
1921	294	-	325	1936	421	-	225
1922	235	155	221	1937	151	-	352
1923	366	317	341	1938	185	-	151
1924	329	273	-	1939	273	-	120

1. Quelle station utiliseriez-vous pour compléter les données manquantes à la station B ?
2. Reporter sur du papier millimétré les k couples de valeurs ;
3. Calculer le coefficient de corrélation ;
4. Déterminer les paramètres a et b de la droite de régression ;
5. Reconstituer les valeurs manquantes de la série des pluies annuelles de la station B,
6. Tracer la droite de régression.

Anonyme / USTHB – FGC – 1MIE-1112 / TD N°2 d'Hydrologie

Solution.

1- La station utilisait pour compléter les données manquantes à la station B.

Pour combler les valeurs manquantes à la station B, il faudrait que le nombre de valeurs concomitantes dans la série à expliquer et la série explicative soit plus grand que 10.

Entre la série A et la série B, il y a 16 valeurs concomitantes, alors qu'il n'y a que 8 entre les séries B et C. Donc, on doit utiliser *la série A pour compléter la série B.*

2- k couples de valeurs reportées sur du papier millimétré

Les années communes figurent dans le tableau II.1 ci-dessous :

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad \text{A. N} \Rightarrow r = \frac{44583}{\sqrt{48458} \sqrt{60835,4}}$$

r = 0,821

$$r = \frac{n \sum_{i=1}^n x_i y_i - (\sum_{i=1}^n x_i)(\sum_{i=1}^n y_i)}{\sqrt{n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2} \sqrt{n \sum_{i=1}^n y_i^2 - (\sum_{i=1}^n y_i)^2}}$$

$$\text{A. N} \Rightarrow r = \frac{(16 \times 1212388) - (4976 \times 3755)}{\sqrt{16 \times 1595994 - (4976)^2} \times \sqrt{16 \times 942087 - 3755^2}}$$

r = 0,821

$$r = \frac{\sum_{i=1}^n x_i y_i - \frac{(\sum_{i=1}^n x_i)(\sum_{i=1}^n y_i)}{n}}{\sqrt{\sum_{i=1}^n x_i^2 - \frac{(\sum_{i=1}^n x_i)^2}{n}} \sqrt{\sum_{i=1}^n y_i^2 - \frac{(\sum_{i=1}^n y_i)^2}{n}}}$$

$$\text{A. N} \Rightarrow r = \frac{1212388 - \frac{4976 \times 3755}{16}}{\sqrt{1595994 - \frac{(4976)^2}{16}} \times \sqrt{942087 - \frac{(3755)^2}{16}}}$$

r = 0,8211

4 Paramètres a et b de la droite de régression

Les paramètres a et b de la droite de régression $y = ax + b$ sont donnés par les relations suivantes ;

$$a = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad \text{A. N} \Rightarrow a = \frac{44583}{48458}$$

a = 0,920

$$a = \frac{n \sum_{i=1}^n x_i y_i - (\sum_{i=1}^n x_i)(\sum_{i=1}^n y_i)}{n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2} \quad \text{A. N} \Rightarrow a = \frac{(16 \times 1212388) - (4976 \times 3755)}{16 \times 1595994 - (4976)^2}$$

a = 0,920

$$a = \frac{\sum_{i=1}^n x_i y_i - \frac{(\sum_{i=1}^n x_i)(\sum_{i=1}^n y_i)}{n}}{\sum_{i=1}^n x_i^2 - \frac{(\sum_{i=1}^n x_i)^2}{n}} \quad \text{A. N} \Rightarrow a = \frac{1212388 - \frac{4976 \times 3755}{16}}{1595994 - \frac{(4976)^2}{16}}$$

a = 0,920

$$a = r \frac{S_y}{S_x} \quad \text{A. N} \Rightarrow a = 0,8211 \times \frac{64,7}{56,6}$$

a = 0,920

$$b = \bar{y} - a\bar{x} \quad \text{A. N} \Rightarrow b = 236,8 - 0,920 \times 310,8$$

a = 51,4

D'où l'équation de la droite de régression : **y = 0,920 x - 51,4. (i)**

5- Reconstitution des valeurs manquantes :

Pour reconstituer les valeurs manquantes de la série B, nous avons appliqué l'équation de la droite de régression ci-dessus (Equation i). Les valeurs reconstituées sont en rouges (tableau II.3).

Tableau II.3. Reconstituer les valeurs manquantes de la série B.

Année	A	B	Année	A	B
1910	222	152,8	1925	266	136
1911	150	86,6	1926	327	184
1912	281	207,1	1927	310	252
1913	210	141,8	1928	360	285
1914	172	106,8	1929	413	318
1915	260	229	1930	258	206
1916	252	208	1931	225	156
1917	321	204	1932	308	237
1918	382	359	1933	-	209
1919	364	236	1934	-	209
1920	193	126,1	1935	349	
1921	294		1936	421	
1922	235	155	1937	151	
1923	366	317	1938	185	
1924	329	273	1939	273	

Nous pouvons combler les valeurs manquantes de la série A en utilisant l'équation de régression B en A.

6- Tracé de la droite de régression

Le tableau II.4 suivant représente les valeurs théoriques données par la droite de régression (Equation i) (Fig. II.4).

Tableau II.4. Valeurs théoriques de la droite de régression.

A	B (Pts expérimentaux)	B' (Droite de régression)	A	B (Pts expérimentaux)	B' (Droite de régression)
222	152,8	152,8	329	273	251,2
150	86,6	86,6	266	136	193,3
281	207,1	207,1	327	184	249,4
210	141,8	141,8	310	252	233,8
172	106,8	106,8	360	285	279,8
260	229	187,8	413	318	328,5
252	208	180,4	258	206	185,9
321	204	243,9	225	156	155,6
382	359	300,0	308	237	231,9
364	236	283,4	349		269,6
193	126,1	126,1	421		335,9
294		219,0	151		87,5
235	155	164,8	185		118,8
366	317	285,3	273		199,7

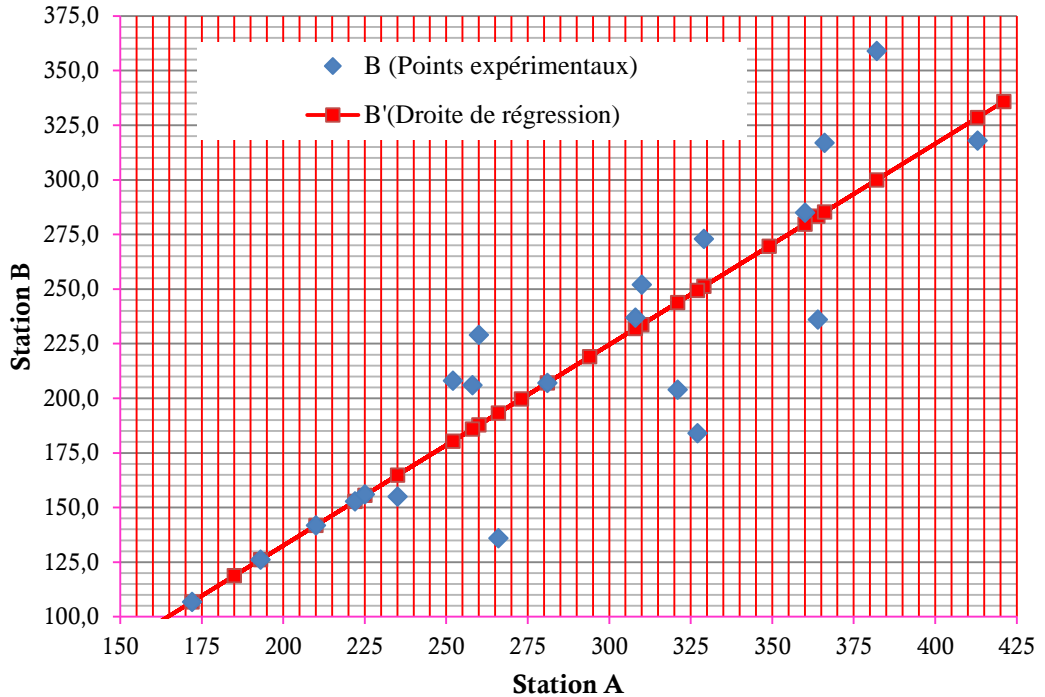


Fig. II.4. Valeurs théoriques données par la droite de régression.

Les nouvelles moyennes et écart-type sont respectivement : 208 mm et 72,82 mm. Les anciennes étaient 236,8 mm et 54,7 mm.

Exercice 2.

Vérifier l'homogénéité de la série des pluies annuelles maximales de la station pluviométrique de Magnhia pour un risque de 5% en utilisant :

- le test de Wilcoxon
- le test de Mann-Whitney

Tableau.

Années	Pluies (mm)	Années	Pluies (mm)	Années	Pluies (mm)
1975-1976	30,2	1987-1988	25,6	1999-2000	26,5
1976-1977	45,7	1988-1989	51,6	2000-2001	32,6
1977-1978	50	1989-1990	33,5	2001-2002	64,7
1978-1979	18,3	1990-1991	32,5	2002-2003	29,4
1979-1980	53	1991-1992	42,5	2003-2004	70,8
1980-1981	55	1992-1993	39	2004-2005	24,5
1981-1982	14,5	1993-1994	25	2005-2006	22,5
1982-1983	15,5	1994-1995	25	2006-2007	26,5
1983-1984	21,1	1995-1996	43,9	2007-2008	30,5
1984-1985	30,4	1996-1997	21,6	2008-2009	27
1985-1986	29,3	1997-1998	28,2	2009-2010	26
1986-1987	80,7	1998-1999	44,8	2010-2011	32

Solution.

1- Test de Wilcoxon

On commence par diviser notre série des données pluviométriques en deux sous-échantillons de tailles respectives N_1 et N_2 avec $N_1 < N_2$.

Dans notre cas, nous prenons $N_1 = 16$ valeurs et $N_2 = 20$ valeurs ($N = N_1 + N_2 = 16 + 20 = 36$). Le tableau II.5 indique les étapes de calcul pour tester l'homogénéité de la série des pluies annuelles maximales de la station de Maghnia pour un risque de 5% en utilisant le test de Wilcoxon.

- *Colonne. 1* : Dates des mesures de la variable ;
- *Colonne. 2* : Données brutes;
- *Colonne. 3* : Échantillon X;
- *Colonne. 4* : Échantillon Y;
- *Colonne. 5 & Colonne 6* : Rangs et valeurs classées de la série originale ;
- *Colonne 7* : Origine de la valeur de la série, c'est à dire on note si elle provient de l'échantillon X ou de l'échantillon Y ;
- *Colonne 8* : Rang de la valeur qui provient de la série X.

Nous calculons ensuite les valeurs de :

$$W_x = \sum \text{Rang } x = 310$$

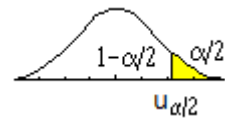
Des deux bornes W_{\min} et W_{\max} , données par les formules suivantes:

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

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

$u_{\frac{1-\alpha}{2}}$: représente la valeur de la variable centrée réduite de la loi normale

correspondant à $\frac{1-\alpha}{2}$ (au seuil de 95 %, nous avons $= 1,96$).



Après calcul, nous avons trouvé que : $W_{\min} = 215,93$ et $W_{\max} = 376,07$

On vérifie l'inégalité: $W_{\min} < W_x = \sum \text{Rang } x < W_{\max}$ c'est-à-dire : $215,93 < 310 < 376,07$

Donc on conclue que notre série des précipitations des pluies maximales de la station de Maghnia est **homogène**.

Tableau II.5. Test de Wilcoxon.

1	2	3	4	5	6	7	8
Années	P(mm)	X	Y	Rang	X ∪ Y	Origine	Rang (X _i)
1975-1976	30,2	30,2	42,5	1	14,5	X	1
1976-1977	45,7	45,7	39	2	15,5	X	2
1977-1978	50	50	25	3	18,3	X	3
1978-1979	18,3	18,3	25	4	21,1	X	4
1979-1980	53	53	43,9	5	21,6	Y	
1980-1981	55	55	21,6	6	22,5	Y	
1981-1982	14,5	14,5	28,2	7	24,5	Y	
1982-1983	15,5	15,5	44,8	8	25	Y	
1983-1984	21,1	21,1	26,5	9	25	Y	
1984-1985	30,4	30,4	32,6	10	25,6	X	10
1985-1986	29,3	29,3	64,7	11	26	Y	
1986-1987	80,7	80,7	29,4	12	26,5	Y	
1987-1988	25,6	25,6	70,8	13	26,5	Y	
1988-1989	51,6	51,6	24,5	14	27	Y	
1989-1990	33,5	33,5	22,5	15	28,2	Y	
1990-1991	32,5	32,5	26,5	16	29,3	X	16
1991-1992	42,5		30,5	17	29,4	Y	
1992-1993	39		27	18	30,2	X	18
1993-1994	25		26	19	30,4	X	19
1994-1995	25		32	20	30,5	Y	
1995-1996	43,9			21	32	Y	
1996-1997	21,6			22	32,5	X	22
1997-1998	28,2			23	32,6	Y	
1998-1999	44,8			24	33,5	X	24
1999-2000	26,5			25	39	Y	
2000-2001	32,6			26	42,5	Y	
2001-2002	64,7			27	43,9	Y	
2002-2003	29,4			28	44,8	Y	
2003-2004	70,8			29	45,7	X	29
2004-2005	24,5			30	50	X	30
2005-2006	22,5			31	51,6	X	31
2006-2007	26,5			32	53	X	32
2007-2008	30,5			33	55	X	33
2008-2009	27			34	64,7	Y	
2009-2010	26			35	70,8	Y	
2010-2011	32			36	80,7	X	36

$$\sum \text{Rang X} = 310$$

2- Test de Man-Whitney

- L'échantillon est divisé en deux sous-ensembles de tailles respectives N_1 et N_2 avec: $N_1 < N_2$.

La taille de l'échantillon original est $N = N_1 + N_2$;

- Les rangs $R(X_i)$ des éléments du premier sous-ensemble N_1 et $R(Y_i)$ ceux des éléments du second sous-ensemble N_2 dans l'échantillon original ;
- K et S comme suit: $K = L - \frac{N_1 \times (N_1 + 1)}{2}$ et $S = N_1 \times N_2 - K$; avec $L = \sum R(X_i)$; c'est à dire la somme des rangs des éléments de l'échantillon N_1 dans l'échantillon original ;
- K est la somme des nombres de dépassements de chaque élément du second échantillon par ceux du premier échantillon ;
- S est la somme des nombres de dépassements des éléments du premier sous-ensemble (ou échantillon) par ceux du second.

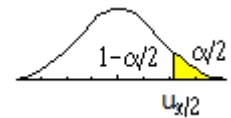
Nous montrons que lorsque $N > 36$, $N_1 > 3$ et $N_2 > 3$; K et S sont distribués selon une loi normale ayant :

- une moyenne égale à: $\bar{K} = \bar{S} = \frac{N_1 \times N_2}{2}$

- et un écart-type égal à: $S_k = S_s = \frac{N_1 \times N_2}{12} \times (N_1 + N_2 + 1)$

On peut alors tester l'hypothèse H_0 que les deux sous-ensembles proviennent de la même population, au niveau de signification α , en comparant la grandeur:

$T = \left| \frac{K - \bar{K}}{s_k} \right|$ avec la variable normale centrée réduite ayant une probabilité de dépassement $\frac{\alpha}{2}$. Si $T < u_{\frac{1-\alpha}{2}}$ on accepte H_0



Le tableau II.6 indique les étapes de calcul pour tester l'homogénéité de la série des pluies annuelles maximales de la station de Maghnia pour un risque de 5% en utilisant le test de Mann-Whitney.

- *Colonne 1* : Dates des mesures de la variable ;
- *Colonne 2* : Données brutes;
- *Colonne 3* : Pluies triées par ordre croissant ;
- *Colonne 4* : Rangs des données triées ;
- *Colonne 5* : Echantillon X ;
- *Colonne 6* : Rang de chaque valeur du sous-ensemble N_1 dans l'échantillon original de 36 valeurs classées ;
- *Colonne 7* : Echantillon Y ;
- *Colonne 8* : Rang de chaque valeur du sous-ensemble N_2 dans l'échantillon original de 36 valeurs classées ;
- *Colonne 9* : Echantillon X triées ;
- *Colonne 10* : Nombre de fois que chaque élément du sous ensemble N_1 est dépassé par les éléments du sous-ensemble N_2 , la somme des éléments de cette colonne est égale à $K = 146$.
- *Colonne 11* : Echantillon X triées ;
- *Colonne 12* : Nombre de fois que chaque élément du sous-ensemble N_2 est dépassé par les éléments du sous-ensemble N_1 , la somme des valeurs de cette colonne est égale à $S = 174$.

Tableau II.6. Test de Mann-Whitney.

1	2	3	4	5	6	7	8	9	10	11	12		
Année	P(mm)	Pclass. (mm)	Rangs	X	Rangs de X	Y	Rangs de Y	X class.	Nbre Dépass.	Y class.	Nbre Dépass.		
1975-1976	30,2	14,5	1	30,2	18	42,5	26	14,5	20	21,6	12		
1976-1977	45,7	15,5	2	45,7	29	39	25	15,5	20	22,5	12		
1977-1978	50	18,3	3	50	30	25	8	18,3	20	24,5	12		
1978-1979	18,3	21,1	4	18,3	3	25	9	21,1	20	25	12		
1979-1980	53	21,6	5	53	32	43,9	27	25,6	15	25	12		
1980-1981	55	22,5	6	55	33	21,6	5	29,3	10	26	11		
1981-1982	14,5	24,5	7	14,5	1	28,2	15	30,2	9	26,5	11		
1982-1983	15,5	25	8	15,5	2	44,8	28	30,4	9	26,5	11		
1983-1984	21,1	25	9	21,1	4	26,5	12	32,5	7	27	11		
1984-1985	30,4	25,6	10	30,4	19	32,6	23	33,5	6	28,2	11		
1985-1986	29,3	26	11	29,3	16	64,7	34	45,7	2	29,4	10		
1986-1987	80,7	26,5	12	80,7	36	29,4	17	50	2	30,5	8		
1987-1988	25,6	26,5	13	25,6	10	70,8	35	51,6	2	32	8		
1988-1989	51,6	27	14	51,6	31	24,5	7	53	2	32,6	7		
1989-1990	33,5	28,2	15	33,5	24	22,5	6	55	2	39	6		
1990-1991	32,5	29,3	16	32,5	22	26,5	13	80,7	0	42,5	6		
1991-1992	42,5	29,4	17			30,5	20			43,9	6		
1992-1993	39	30,2	18			27	14			44,8	6		
1993-1994	25	30,4	19			26	11			64,7	1		
1994-1995	25	30,5	20			32	21			70,8	1		
1995-1996	43,9	32	21										
1996-1997	21,6	32,5	22										
1997-1998	28,2	32,6	23										
1998-1999	44,8	33,5	24										
1999-2000	26,5	39	25										
2000-2001	32,6	42,5	26										
2001-2002	64,7	43,9	27										
2002-2003	29,4	44,8	28										
2003-2004	70,8	45,7	29										
2004-2005	24,5	50	30										
2005-2006	22,5	51,6	31										
2006-2007	26,5	53	32										
2007-2008	30,5	55	33										
2008-2009	27	64,7	34										
2009-2010	26	70,8	35										
2010-2011	32	80,7	36										
				L=	310					K=	146	S=	174

D'après les calculs : L = 310, K = 146 et S = 174 ;

$$\bar{K} = \bar{S} = \frac{N_1 \times N_2}{2} = \frac{16 \times 20}{2} = 160$$

$$S_k = S_s = \frac{N_1 \times N_2}{12} \times (N_1 + N_2 + 1) = \frac{16 \times 20}{12} \times (16 + 20 + 1) = 986,67$$

$$T = \left| \frac{K - \bar{K}}{s_k} \right| = \left| \frac{146 - 160}{986,67} \right| = 0,0142$$

Pour $\alpha = 95\%$ on a $u_{\frac{1-\alpha}{2}} = 1,96 > T = 0,0142$

Ce qui veut dire qu'on peut accepter l'hypothèse H_0 que les deux sous-ensembles proviennent de la même population et que notre série des totaux pluviométriques maximaux à l'échelle annuelle de la station de Maghnia est **homogène**.

Exercice 3.

Vérifier l'homogénéité des séries de précipitations annuelles des stations P1 et P2 en utilisant la station de référence par:

- le test de test de Student pour $\alpha = 5\%$,
- la méthode du double cumul,
- la méthode des cumuls des résidus avec un intervalle de confiance de 99%.

Tableau.

Année	Station de référence (mm)	Station P1 (mm)	Station P2 (mm)	Année	Station de référence (mm)	Station P1 (mm)	Station P2 (mm)
1971	1125	1050	1275	1981	1360	1132	1195
1972	1208	1165	1356	1982	1269	1155	1120
1973	1105	1021	1242	1983	1218	1065	1063
1974	1088	1041	1250	1984	1145	1035	1051
1975	1175	1100	1300	1985	964	1070	1072
1976	1025	1012	1150	1986	1235	1263	1265
1977	1080	1037	1166	1987	766	666	663
1978	1100	1022	1292	1988	931	915	918
1979	1021	1014	1135	1989	912	906	902
1980	895	950	930	1990	806	763	764

Sari, 2002b

Solution.

1- Test de Student

Soit X une variable normale connue sur deux échantillons de taille N_1 et N_2 soit \bar{X}_1 , \bar{X}_2 et S_1 , S_2 , les moyennes et écart-types calculés sur les deux échantillons 1 et 2.

Soit

$$t = \frac{\bar{X}_1 - \bar{X}_2}{S \sqrt{\frac{1}{N_1} + \frac{1}{N_2}}}$$

Ou ;

$$S = \sqrt{\frac{(N_1 - 1)S_1^2 + (N_2 - 1)S_2^2}{N_1 + N_2 - 2}}$$

La variable **t** suit une *loi de probabilité de Student*, dont le paramètre (appelé degré de liberté) vaut $N_1 + N_2 - 2$.

Selon la valeur de t et donc de la probabilité d'apparition d'une telle valeur, on décide s'il est plausible ou pas de considérer que les deux moyennes \bar{X}_1 et \bar{X}_2 deux estimations de celle de la population μ (homogénéité des moyennes).

Si $t(N_1 + N_2 - 2) > t_{\text{calculé}}$ pour un seuil de signification $(1 - \alpha)$, les deux échantillons sont considérés comme **homogènes**.

Dans notre cas, nous allons comparer la station de référence successivement avec les stations P_1 et P_2 . Nous avons :

$$S_{\text{Ref},P_1} = \sqrt{\frac{(N_1 - 1)S_1^2 + (N_2 - 1)S_2^2}{N_1 + N_2 - 2}} \quad \text{A. N} \Rightarrow S_{\text{Ref},P_1} = \sqrt{\frac{(20 - 1)(156,7)^2 + (20 - 1)(134,6)^2}{20 + 20 - 2}}$$

$$\Rightarrow S_{\text{Ref},P_1} = \mathbf{689,78}$$

$$S_{\text{Ref},P_2} = \sqrt{\frac{(N_1 - 1)S_1^2 + (N_3 - 1)S_3^2}{N_1 + N_3 - 2}} \quad \text{A. N} \Rightarrow S_{\text{Ref},P_2} = \sqrt{\frac{(20 - 1)(156,7)^2 + (20 - 1)(187,7)^2}{20 + 20 - 2}}$$

$$\Rightarrow S_{\text{Ref},P_2} = \mathbf{695,96}$$

Avec, N_1 , N_2 et N_3 sont les longueurs des séries, \bar{X}_1 , \bar{X}_2 et \bar{X}_3 leurs moyennes et S_1 , S_2 et S_3 leurs écarts types. D'où

$$t_{1,2} = \frac{\bar{X}_1 - \bar{X}_2}{S_{1,2} \times \sqrt{\frac{1}{N_1} + \frac{1}{N_2}}} \quad \text{A. N} \Rightarrow t_{1,2} = \frac{1071,4 - 1019,1}{689,78 \times \sqrt{\frac{1}{20} + \frac{1}{20}}}$$

$$t_{1,2} = \mathbf{0,240}$$

Et,

$$t_{1,3} = \frac{\bar{X}_1 - \bar{X}_3}{S_{1,2} \times \sqrt{\frac{1}{N_1} + \frac{1}{N_3}}} \quad \text{A. N} \Rightarrow t_{1,3} = \frac{1071,4 - 1105,45}{689,78 \times \sqrt{\frac{1}{20} + \frac{1}{20}}}$$

$$t_{1,3} = \mathbf{-0,156}$$

Le nombre de degrés de liberté est $N_1 + N_3 - 2 = 38$

Le seuil de signification est $(1 - \alpha) = (1 - 0,05) = 0,95$.

Pour ces valeurs la table de Student donne $t = 1,686 > t_{1,2} = 0,240$ et $t = 1,686 > t_{1,2} = -0,156$; ce qui veut dire que **les deux séries sont homogènes**.

2- Méthode des doubles cumuls

Elle permet de détecter la non-homogénéité d'une série de mesures et de la corriger. La méthode consiste à comparer les pluies (ou toute autre variable) cumulées d'une station B, à propos de laquelle on éprouve des doutes quant à son homogénéité, avec les pluies cumulées d'une station A dont les mesures sont jugées homogènes.

Application de la méthode à la série P1

On commence par établir le tableau II.7 ci-dessous :

- *Colonne. 1* : Dates des mesures de la variable ;
- *Colonne. 2* : Précipitations mesurées à la station A. Station de référence (mm) ;
- *Colonne. 3* : Précipitations mesurées à la station B. Station à tester (mm) ;
- *Colonne. 4* : Cumuls des pluies à la station A (mm) ;
- *Colonne. 5* : Cumuls des pluies à la station B (mm).

Ensuite, les valeurs sont portées sur du papier millimétré, avec les valeurs cumulées de A en abscisses et les valeurs de B en ordonnées.

Tableau II.7. *Méthode de double cumul. A station de référence et B Station à tester*

1	2	3	4	5
Année	Station de référence A (mm)	Station B P1 (mm)	Cumuls A	Cumuls B
1990	806	763	806	763
1989	912	906	1718	1669
1988	931	915	2649	2584
1987	766	666	3415	3250
1986	1235	1263	4650	4513
1985	964	1070	5614	5583
1984	1145	1035	6759	6618
1983	1218	1065	7977	7683
1982	1269	1155	9246	8838
1981	1360	1132	10606	9970
1980	895	950	11501	10920
1979	1021	1014	12522	11934
1978	1100	1022	13622	12956
1977	1080	1037	14702	13993
1976	1025	1012	15727	15005
1975	1175	1100	16902	16105
1974	1088	1041	17990	17146
1973	1105	1021	19095	18167
1972	1208	1165	20303	19332
1971	1125	1050	21428	20382

Graphiquement (Fig. II.5) que les points s'alignent sur un seul segment de droite, ce qui est interprété comme quoi la série B (P1) est homogène.

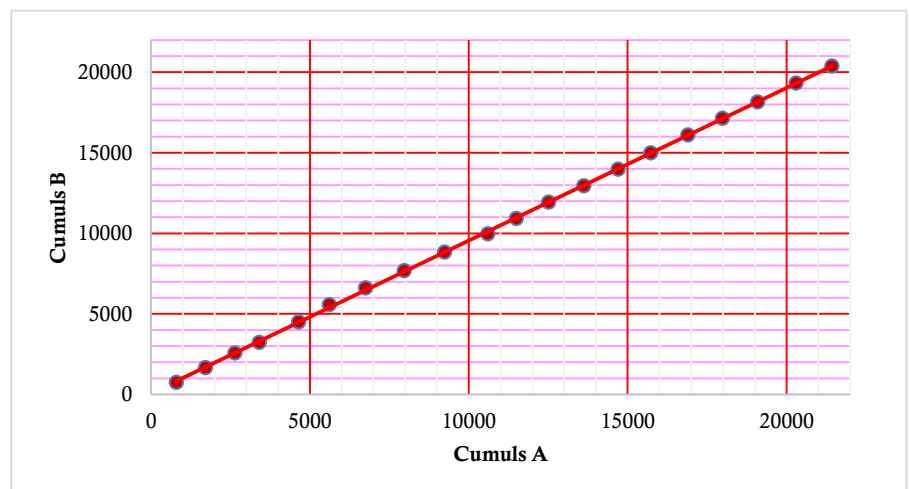


Fig. II.5. *Méthode de double cumul. A station de référence et B Station à tester.*

Application de la méthode à la série P2

On commence par établir le tableau II.8 ci-dessous :

- *Colonne. 1* : Dates des mesures de la variable ;
- *Colonne. 2* : Précipitations mesurées à la station A. Station de référence (mm) ;
- *Colonne. 3* : Précipitations mesurées à la station C. Station à tester (mm) ;
- *Colonne. 4* : Cumuls des pluies à la station A (mm) ;
- *Colonne. 5* : Cumuls des pluies à la station C (mm).

Ensuite, les valeurs sont portées sur du papier millimétré (Fig. II.6), avec les valeurs cumulées de A en abscisses et les valeurs de C en ordonnées.

Tableau II.8. *Méthode de double cumul. A station de référence et C Station à tester.*

1	2	3	4	5
Année	Station de référence A (mm)	Station C P2 (mm)	Cumuls A	Cumuls C
1990	806	764	806	764
1989	912	902	1718	1666
1988	931	918	2649	2584
1987	766	663	3415	3247
1986	1235	1265	4650	4512
1985	964	1072	5614	5584
1984	1145	1051	6759	6635
1983	1218	1063	7977	7698
1982	1269	1120	9246	8818
1981	1360	1195	10606	10013
1980	895	930	11501	10943
1979	1021	1135	12522	12078
1978	1100	1292	13622	13370
1977	1080	1166	14702	14536
1976	1025	1150	15727	15686
1975	1175	1300	16902	16986
1974	1088	1250	17990	18236
1973	1105	1242	19095	19478
1972	1208	1356	20303	20834
1971	1125	1275	21428	22109

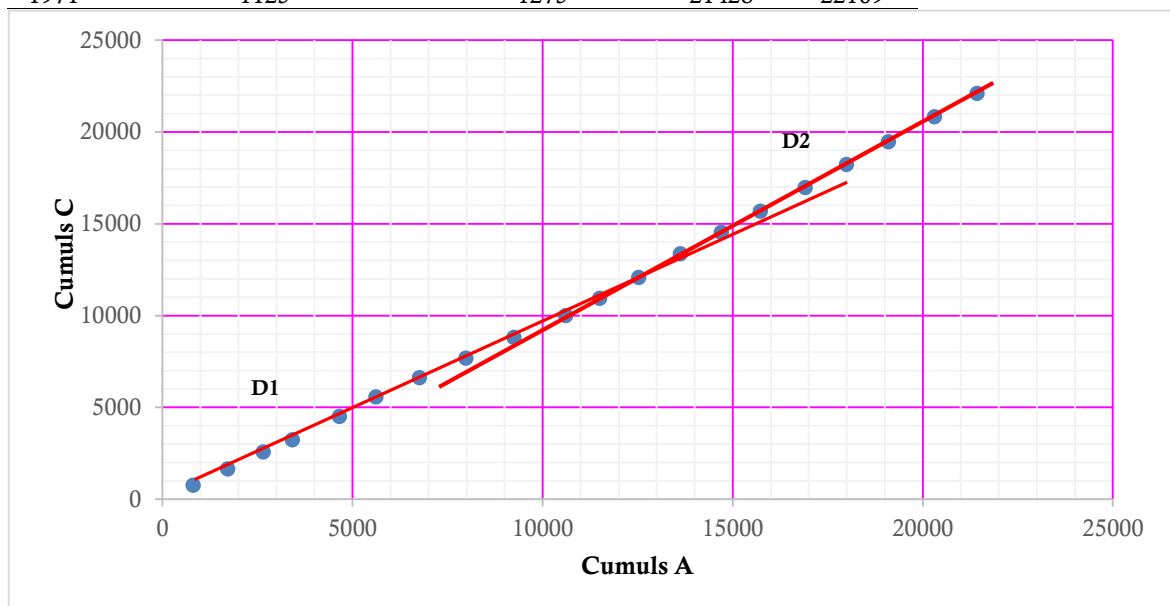


Fig. II.6. *Méthode de double cumul. A station de référence et C Station à tester.*

On voit sur le graphique que les points s'alignent sur deux segments de droite différents, c'est-à-dire qu'il y a une cassure sur la droite au cours de l'année 1979. On suppose que le déplacement (ou autre cause d'erreur) s'est produit en 1979. Les données mesurées après 1979 sont jugées bonnes et on ne doit corriger que les données précédentes (1979 à 1971).

La décision de corriger ou non les données de l'année 1979 est prise après une connaissance détaillée des circonstances de « l'accident » au cours de cette année.

On calcule les pentes m_1 du segment de droite qui contient les données de 1990 à 1979, et m_2 du segment de droite qui contient les données de 1979 à 1971.

$$m_1 = \frac{22109 - 12078}{21428 - 12522} \Rightarrow m_1 = \frac{10\,031}{8\,906}$$

$$m_1 = \mathbf{1,1263}$$

$$m_2 = \frac{12\,078 - 764}{12522 - 806} \Rightarrow m_2 = \frac{11\,314}{11\,716}$$

$$m_2 = \mathbf{0,9657}$$

On calcule le rapport des pentes $\frac{m_1}{m_2}$ avec lequel on va multiplier les données des années 1979 à 1971 pour les corriger.

$$\frac{m_2}{m_1} = \frac{0,9657}{1,1263} = \mathbf{0,765}$$

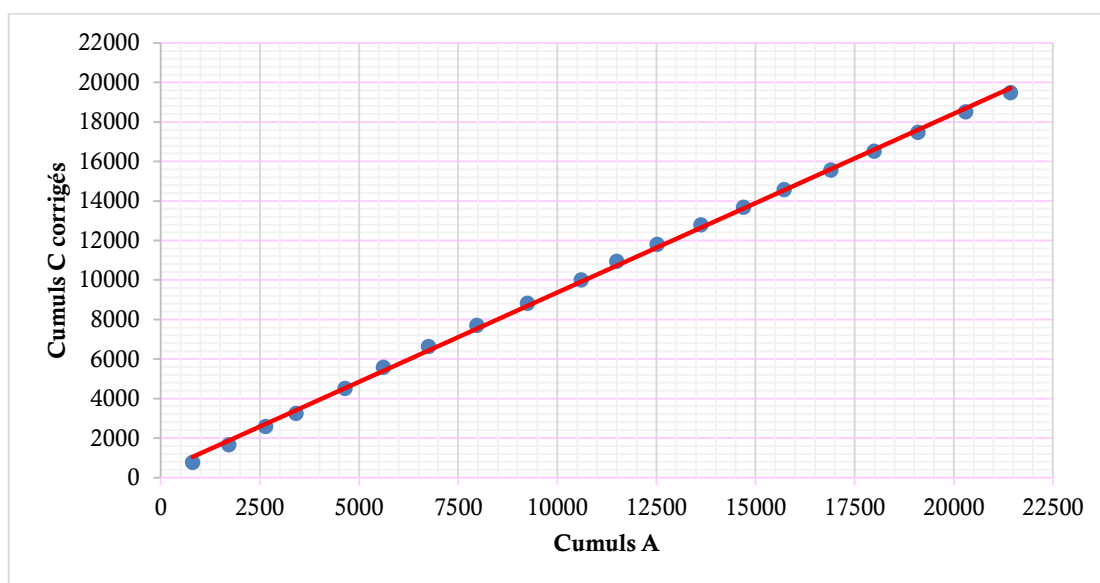
Pour corriger la station C, nous établirons le tableau II.9 :

- *Colonne. 6* : Valeurs des pluies corrigées à la station C (mm) ;
- *Colonne. 7* : Cumuls des pluies à la station C (mm).

Une fois ces données corrigées, on refait l'opération (Fig. II.7).

Tableau II.9. Méthode de double cumul. A station de référence et C Station à tester corrigée.

1	2	3	4	5	6	7
Année	Station de référence A (mm)	Station C P2 (mm)	Cumuls A	Cumuls C	C corrigés	Cumuls C corrigés
1990	806	764	806	764	764	764
1989	912	902	1718	1666	902	1666
1988	931	918	2649	2584	918	2584
1987	766	663	3415	3247	663	3247
1986	1235	1265	4650	4512	1265	4512
1985	964	1072	5614	5584	1072	5584
1984	1145	1051	6759	6635	1051	6635
1983	1218	1063	7977	7698	1063	7698
1982	1269	1120	9246	8818	1120	8818
1981	1360	1195	10606	10013	1195	10013
1980	895	930	11501	10943	930	10943
1979	1021	1135	12522	12078	868,275	11811
1978	1100	1292	13622	13370	988,38	12800
1977	1080	1166	14702	14536	891,99	13692
1976	1025	1150	15727	15686	879,75	14571
1975	1175	1300	16902	16986	994,5	15566
1974	1088	1250	17990	18236	956,25	16522
1973	1105	1242	19095	19478	950,13	17472
1972	1208	1356	20303	20834	1037,34	18510
1971	1125	1275	21428	22109	975,375	19485

**Fig. II.7.** Méthode de double cumul. A station de référence et C Station à tester corrigée.

L'on voit que les points s'alignent sur une droite sans cassure; notre série a donc été rendue homogène. Si l'on constate une autre cassure, on recommence l'opération.

3- Méthode des cumuls des résidus

On appelle résidu ε_i la différence entre la valeur observée d'une variable et la valeur de la même variable obtenue à partir de l'équation de la droite de régression entre la variable explicative et la variable à expliquer.

Dans cet exercice, nous avons une série de référence A et deux séries pluviométriques P1 et P2 dont on veut en connaître l'homogénéité. On calcule les caractéristiques suivantes :

	A	P1	P2
Moyenne (mm)	1071,4	1019,1	1105,5
Ecart type (mm)	156,7	134,6	187,7

Entre la série A et la série P1 $\Rightarrow \hat{P}1_i = 0,758 \times A_i + 207,4$ et $r_{AP1} = 0,883$.

Entre la série A et la série P2 $\Rightarrow \hat{P}2_i = 0,935 \times A_i + 103,7$ et $r_{AP2} = 0,883$.

On a donc, $\varepsilon_i = P_i - \hat{P}_i$.

Le cumul des k premiers résidus est appelé $Z_k = \varepsilon_1 + \varepsilon_2 + \varepsilon_3 + \dots \dots \varepsilon_k$ pour $k < n + 1$.

On sait que :

- La moyenne des $\varepsilon_i = 0$, d'où $Z_n = \sum_1^n \varepsilon_i = 0$

- L'écart type des $\varepsilon_i = \sqrt{S_p \times (1 - r_{AP}^2)}$.

On trace sur du papier graphique Z_k en fonction de k, on obtient le tracé des cumuls des résidus qui sera une courbe oscillant autour de l'axe des X. Ensuite, on définit pour chaque valeur de k l'écart ou résidu à ne pas dépasser pour un seuil de confiance α donné. Cet écart est représenté par la longueur du segment ayant pour abscisse k centré sur l'axe des X et de longueur égale à : $\frac{u_{1-\alpha}}{2} \times \text{écart type de } Z_k$ où $\frac{u_{1-\alpha}}{2}$ est la variable centrée réduite de Gauss ayant une probabilité de dépassement égale à $\frac{1-\alpha}{2}$ et l'écart type de $Z_k = S_p^2 \times (1 - r_{AP}^2) \times k \times (n - k) \times (n - 1) / n^2$.

Les extrémités des segments d'abscisse k et centrés sur l'axe des x décrivent une ellipse de grand axe égal à n. Ceci veut dire que pour une valeur donnée de k, il y a une probabilité α pour que le point représentatif de Z_k soit à l'intérieur de ce segment. Mais cela ne veut pas dire que pour l'ensemble du tracé « la probabilité » qu'aucun point ne sorte soit égale à α (où $\alpha = \text{seuil de signification} = 0,99 \rightarrow \frac{u_{1-\alpha}}{2} = 2,57$).

3-1 Méthode des cumuls des résidus appliquée à la série P1

Le tableau II.10 ci-dessous résume les calculs. Ces résultats de calcul sont représentés dans la figure II.8 :

- *Colonne 1* : Différentes valeurs de k ;
- *Colonne 2* : Années de mesure ;
- *Colonne 3* : Valeurs des pluies mesurées en mm, à la station de référence A ;
- *Colonne 4* : Valeurs des pluies mesurées en mm, à la station P1 ;

- Colonne 5 : Valeurs des pluies théoriques à la station P1 obtenues grâce à l'équation de régression :

$$\hat{P}_i = 0,758 \times A_i + 207,4$$

- Colonne 6 : Calculé les résidus ε_i grâce à l'équation : $\varepsilon_i = P1_i - \hat{P}_i$

- Colonne 7 : Cumuls des résidus : $Cumul_k = Cumul_{k-1} + \varepsilon_k$

- Colonne 8 : Haut de l'ellipse donné par l'équation :

$$u_{\frac{1-\alpha}{2}} \times \text{écart type de } Z_k = u_{\frac{1-\alpha}{2}} \times S_{P1}^2 \times (1 - r_{AP1}^2) \times k \times (n - k) \times \frac{n-1}{n^2}$$

$$u_{\frac{1-\alpha}{2}} \times \text{écart type de } Z_k = 2,57 \times 134,6^2 \times (1 - 0,883^2) \times k \times (24 - k) \times 23/24^2$$

- Colonne 9 : Bas de l'ellipse donné par l'équation :

$$u_{\frac{1-\alpha}{2}} \times \text{écart type de } Z_k = u_{\frac{1-\alpha}{2}} \times S_{P1}^2 \times (1 - r_{AP1}^2) \times k \times (n - k) \times \frac{n-1}{n^2}$$

$$u_{\frac{1-\alpha}{2}} \times \text{écart type de } Z_k = -2,57 \times 134,6^2 \times (1 - 0,883^2) \times k \times (20 - k) \times 19/20^2$$

Tableau II.10. Méthode des cumuls des résidus. A station de référence et P1 Station à tester.

1	2	3	4	5	6	7	8	9
K	Année	Station de référence (mm)	Station P1 (mm)	P1 théorique	Résidus	Cumuls des Résidus	Haut de l'ellipse	Bas de l'ellipse
0					0	0	0,0	0,0
1	1990	806	763	818,02	-55,02	-55,02	69,0	-69,0
2	1989	912	906	898,33	7,67	-47,35	130,8	-130,8
3	1988	931	915	912,73	2,27	-45,08	185,3	-185,3
4	1987	766	666	787,71	-121,71	-166,79	232,5	-232,5
5	1986	1235	1263	1143,05	119,95	-46,84	272,4	-272,4
6	1985	964	1070	937,73	132,27	85,43	305,1	-305,1
7	1984	1145	1035	1074,86	-39,86	45,57	330,5	-330,5
8	1983	1218	1065	1130,17	-65,17	-19,60	348,7	-348,7
9	1982	1269	1155	1168,81	-13,81	-33,42	359,6	-359,6
10	1981	1360	1132	1237,76	-105,76	-139,17	363,2	-363,2
11	1980	895	950	885,45	64,55	-74,62	359,6	-359,6
12	1979	1021	1014	980,91	33,09	-41,54	348,7	-348,7
13	1978	1100	1022	1040,77	-18,77	-60,31	330,5	-330,5
14	1977	1080	1037	1025,62	11,38	-48,92	305,1	-305,1
15	1976	1025	1012	983,94	28,06	-20,87	272,4	-272,4
16	1975	1175	1100	1097,59	2,41	-18,46	232,5	-232,5
17	1974	1088	1041	1031,68	9,32	-9,14	185,3	-185,3
18	1973	1105	1021	1044,56	-23,56	-32,69	130,8	-130,8
19	1972	1208	1165	1122,60	42,40	9,71	69,0	-69,0
20	1971	1125	1050	1059,71	-9,71	0,00	0,0	0,0
Som. =	21428	20382						
Moy. =	1071,4	1019,1						
Ecart type =	156,7	134,6						

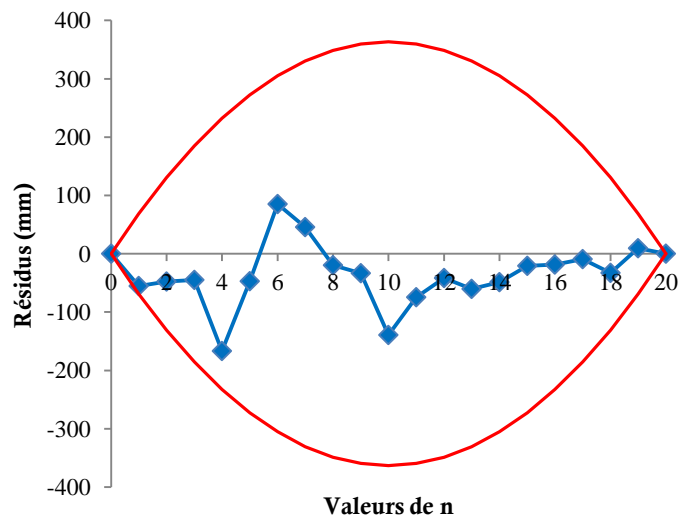


Fig. II.8. Méthode des cumuls des résidus. A station de référence et P1 Station à tester.

3-2 Méthode des cumuls des résidus appliquée à la série P2

Le tableau 3.5 ci-dessous résume les calculs. Ces résultats de calcul sont représentés dans la figure II.9.

Colonne 1 : Différentes valeurs de k.

Colonne 2 : Années de mesure.

Colonne 3 : Valeurs des pluies mesurées en mm, à la station de référence A.

Colonne 4 : Valeurs des pluies mesurées en mm, à la station P2.

Colonne 5 : Valeurs des pluies théoriques à la station P2 obtenues grâce à l'équation de régression :

$$\hat{P}_i = 0,935 \times A_i + 103,7$$

Colonne 6 : Résidus ε_i grâce à l'équation : $\varepsilon_i = P2_i - \hat{P}_i$.

Colonne 7 : Cumuls des résidus : $\text{Cumul}_k = \text{Cumul}_{k-1} + \varepsilon_k$

Colonne 8 : Haut de l'ellipse donné par l'équation :

$$\frac{u_{1-\alpha}}{2} \times \text{écart type de } Z_k = \frac{u_{1-\alpha}}{2} \times S_{P2}^2 \times (1 - r_{AP2}^2) \times k \times (n - k) \times (n - 1) / n^2$$

$$\frac{u_{1-\alpha}}{2} \times \text{écart type de } Z_k = 2,57 \times 187,7^2 \times (1 - 0,781^2) \times k \times (24 - k) \times 23 / 24^2$$

Colonne 9 : Bas de l'ellipse donné par l'équation :

$$-\frac{u_{1-\alpha}}{2} \times \text{écart type de } Z_k = -\frac{u_{1-\alpha}}{2} \times S_{P2}^2 \times (1 - r_{AP2}^2) \times k \times (n - k) \times (n - 1) / n^2$$

$$-\frac{u_{1-\alpha}}{2} \times \text{écart type de } Z_k = -2,57 \times 187,7^2 \times (1 - 0,781^2) \times k \times (20 - k) \times 19 / 20^2$$

Tableau II.11. Méthode des cumuls des résidus. A station de référence et P2 Station à tester.

1	2	3	4	5	6	7	8	9
K	Année	Station de référence (mm)	Station P2 (mm)	P2 théorique	Résidus	Cumuls des Résidus	Haut de l'ellipse	Bas de l'ellipse
0					0	0	0,0	0,0
1	1990	806	764	857,3	-93,31	-93,31	142,0	-142,0
2	1989	912	902	956,4	-54,42	-147,73	269,0	-269,0
3	1988	931	918	974,2	-56,18	-203,92	381,1	-381,1
4	1987	766	663	819,9	-156,92	-360,83	478,3	-478,3
5	1986	1235	1265	1258,4	6,59	-354,24	560,5	-560,5
6	1985	964	1072	1005,0	66,96	-287,28	627,7	-627,7
7	1984	1145	1051	1174,3	-123,26	-410,54	680,0	-680,0
8	1983	1218	1063	1242,5	-179,51	-590,05	717,4	-717,4
9	1982	1269	1120	1290,2	-170,20	-760,25	739,8	-739,8
10	1981	1360	1195	1375,3	-180,28	-940,53	747,3	-747,3
11	1980	895	930	940,5	-10,52	-951,05	739,8	-739,8
12	1979	1021	1135	1058,3	76,67	-874,38	717,4	-717,4
13	1978	1100	1292	1132,2	159,81	-714,57	680,0	-680,0
14	1977	1080	1166	1113,5	52,51	-662,06	627,7	-627,7
15	1976	1025	1150	1062,1	87,93	-574,13	560,5	-560,5
16	1975	1175	1300	1202,3	97,69	-476,44	478,3	-478,3
17	1974	1088	1250	1121,0	129,03	-347,41	381,1	-381,1
18	1973	1105	1242	1136,9	105,14	-242,27	269,0	-269,0
19	1972	1208	1356	1233,2	122,84	-119,44	142,0	-142,0
20	1971	1125	1275	1155,6	119,44	0,00	0,0	0,0
Som. =		21428	22109					
Moy. =		1071,4	1105,45					
Ecart type =		156,7	187,7					

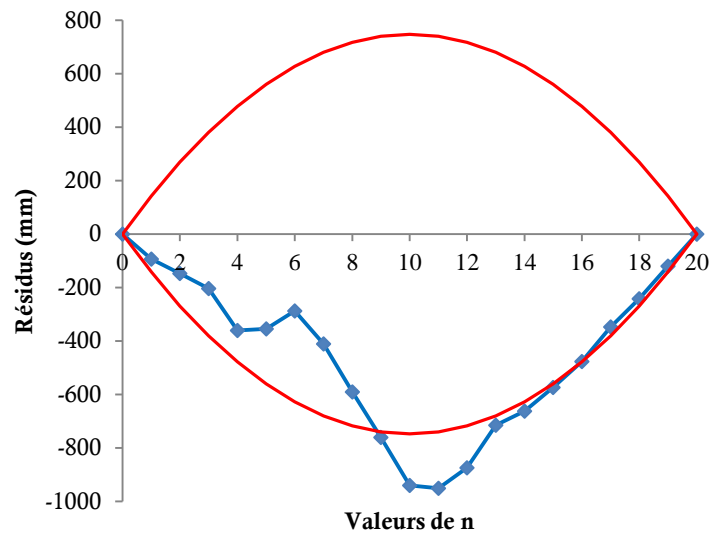
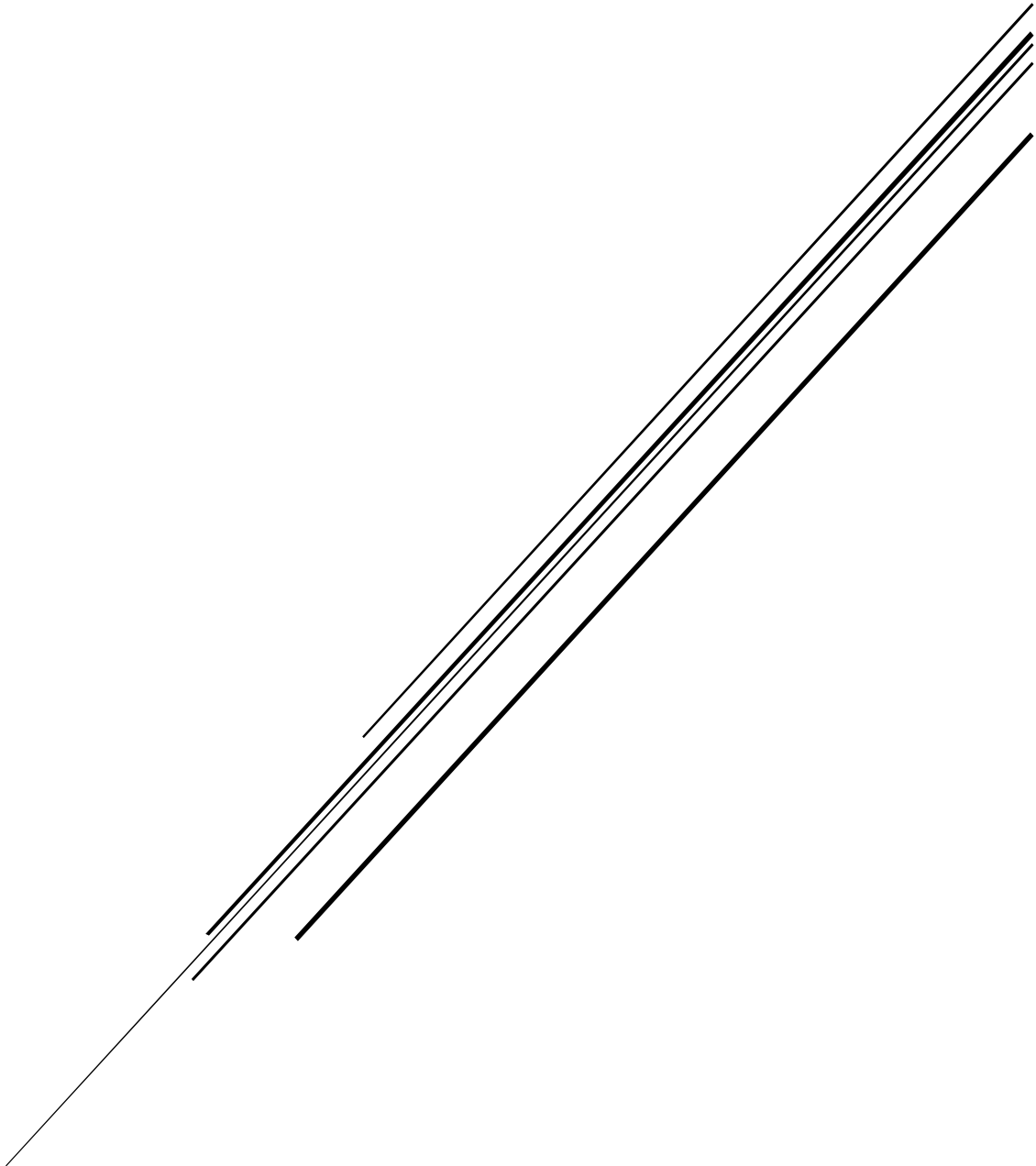


Fig. II.9. *Méthode des cumuls des résidus. A station de référence et P2 Station à tester.*

MODELISATION HYDROLOGIQUE

Chapitre III



MODELISATION HYDROLOGIQUE

En hydrologie le modèle est une nécessité qui tient de la nature même des phénomènes étudiés et des paramètres qui caractérisent leur complexité, leur variabilité spatiale et temporelle. Il s'agit d'une représentation volontairement simplifiée, pour être commode d'emploi, d'un phénomène ou d'un ensemble de phénomènes. Cette représentation pourra être mathématique ou physique, numérique ou analogique.

III.1. Modélisation hydrologique

La modélisation hydrologique est l'application d'expressions mathématiques et logiques qui définissent les relations quantitatives entre les caractéristiques d'un écoulement (sortie) et les facteurs influant sur ses valeurs (entrées). Et parce que les pluies constituent les variables prépondérantes des modèles hydrologiques, ceux-ci sont souvent tout simplement désignés par modèle *pluie-débit* (Fig. III.1).

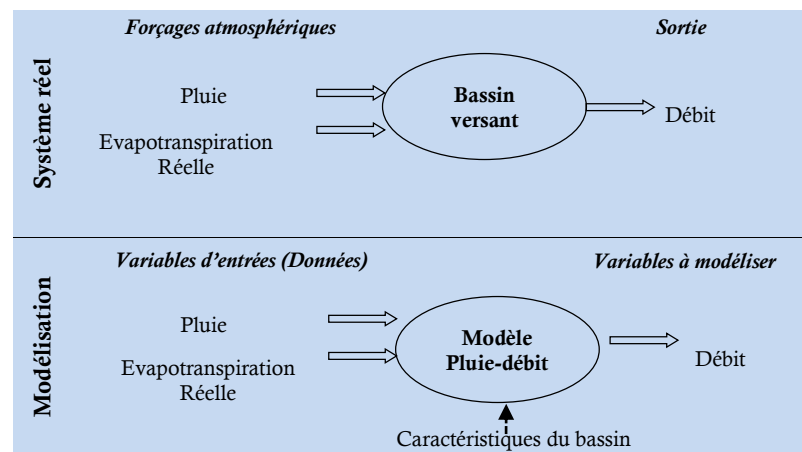


Fig. III.1. Représentation schématique de la modélisation pluie-débit (Gérard, 2010).

Il est essentiel de noter ici, qu'un modèle hydrologique n'a pas pour but de serrer au plus près la réalité des phénomènes qui se déroulent dans un bassin versant, mais de constituer un outil efficace pour résoudre des problèmes pratiques sur les ressources en eau, telle qu'elle est perçue par le modélisateur. En effet, la complexité infinie de la réalité, d'une part, et les restrictions d'ordre pratique imposées par les questions auxquelles le modèle doit donner des réponses, d'autre part, font qu'un modèle sera toujours une représentation incomplète de la réalité. Le

modèle est donc une image de la réalité modélisée par oubli de nombreuses propriétés jugées non pertinentes pour les questions posées (Roche, 1988).

La recherche d'un modèle représentant le fonctionnement d'un système (bassin versant) est connu sous la définition de *problème d'identification*. Résoudre le problème d'identification signifie trouver un *algorithme* (une structure formelle) permettant de décrire les lois de comportement supposées régir le fonctionnement du système (Fig. III.2). On appelle cette phase *codage* (Rosen, 1991 ; Banks, 1998). Il s'agit de simuler l'évolution spatio-temporelle de certaines *variables pronostiques* ou *de sortie* (e.g. le ruissellement) en réponse à l'évolution spatio-temporelle de certaines *variables de forçage* ou *d'entrée* (e.g. les précipitation), en tenant compte, éventuellement, de l'évolution d'un certain nombre de *variable d'état* ou *variables internes* (e.g. l'état d'humidité du bassin) (Zin, 2002).

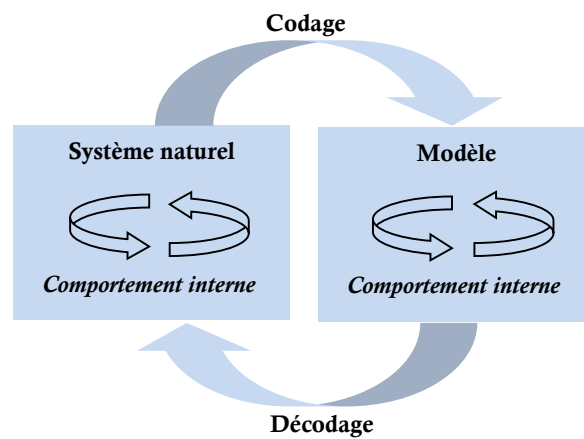


Fig. III.2. Schématisation du processus d'identification d'un modèle (Gupta et Sorooshian., 1985 ; Rosen, 1991).

Des informations caractéristiques du système étudié sont généralement prises en compte dans les modèles. Il peut s'agir de caractéristiques mesurables (avec une certaine précision) du système, et donc directement utilisables en tant que variables d'entrée du modèle. D'autres caractéristiques spécifiques du système peuvent avoir une nature plus subjective, car relatives à la description formelle que l'on fait du système étudié. Elles constituent donc une projection mathématique de notre connaissance, et sont dépendantes des échelles spatio-temporelles de description du système. Ce sont les *paramètres* du modèle (Le Lay, 2006) (Fig. III.3).

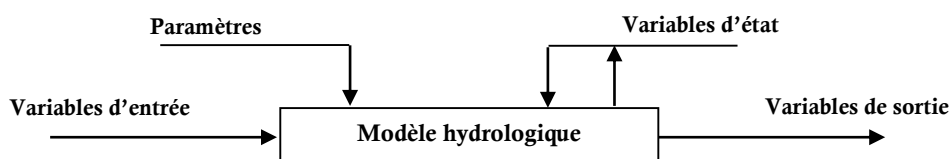


Fig. III.3. Variables d'un modèle hydrologique (Gaume, 2002).

Le développement de ce dernier repose sur trois éléments:

- Le système observé et sa discrétisation spatiale et temporelle, qui en définissent l'objet et ses limites (spatiales ou temporelles). La connaissance du système est conditionnée par la mesure de ses caractéristiques et l'acquisition de données sur les flux, les stocks et les transformations de phase. Dans notre cas, le système est typiquement le *bassin versant*;
- Le modèle est construit pour répondre à des questions et peut ainsi, au-delà d'un outil de représentation, être également un instrument de connaissance. Nous nous intéresserons plus particulièrement à la représentation de la *transformation de la pluie en débit*;
- Le choix d'une formulation de la réalité. L'essentiel de la démarche de modélisation consiste à trouver la formulation de la réalité la plus satisfaisante relativement aux objectifs fixés. En fait, le modèle est la plupart du temps le résultat d'un compromis entre généralité, réalisme et précision (*Kauark-Leite et Nascimento, 1993*). La formulation d'un modèle est conditionnée par la connaissance antérieure des processus ou des systèmes considérés, par les idées et l'imagination du modélisateur et par les hypothèses qui sont formulées.

D'après Thom (1979), il est possible de construire un nombre infini de modèles sur une réalité particulière. Ce sont les questions posées a priori qui sont les balises capables de retreindre le choix du modélisateur. Il souligne qu'un modèle n'est utile que lorsqu'il donne des réponses satisfaisantes aux questions se trouvant à l'origine de son développement.

III.2. Critères de qualité d'un modèle

Lors de l'élaboration d'un modèle, l'ambition de l'hydrologue est d'obtenir un outil le plus performant possible. Il est de coutume d'apprécier la qualité d'un modèle et ses aptitudes selon certains critères tels que la *robustesse* et la *sensibilité*. On souhaite en effet que le modèle ne diverge pas lorsqu'on modifie légèrement la valeur de ses facteurs d'entrée à cause d'erreurs sur les entrées (*robustesse*), mais qu'il soit sensible aux variations significatives des facteurs dont on cherche à simuler les effets, *e.g.* un changement d'occupation des sols (*sensibilité*) (*Chkir, 1994 ; Zin, 2002*). De même, le *principe de parcimonie* doit être respecté : Parmi les différents modèles pouvant répondre à certains objectifs et hypothèses de travail, *il est généralement conseillé de choisir le modèle le plus simple*. Cela revient à choisir le modèle comportant le plus petit nombre possible de paramètres (et de relations non linéaires), car il s'agit de quantités non mesurables. Andersson (1992), ainsi qu'Ambroise (1999), rappellent que la complexité d'un système de modélisation ne doit pas être confondue avec la précision de ses résultats. Il s'agit d'un paradoxe bien connu par les hydrologues : si l'on veut représenter plus finement le fonctionnement d'un système, il est nécessaire d'augmenter l'information qu'il faut fournir au modèle, notamment pour la spécification des valeurs de ses paramètres. Etant donné que les observations sont généralement

limitées, quand on augmente la complexité du modèle, on augmente aussi ses degrés de liberté, diminuant ainsi sa robustesse et sa performance (*Jakeman et Hornberger, 1993*).

Bergström (1991), Lindström *et al.* (1997) ont discuté la notion de complexité optimale du modèle au regard des données disponibles et des objectifs du modèle. Dix-sept ans plus tard, Fenicia *et al.* (2008) rappellent encore qu’augmenter la complexité du modèle n’améliore pas nécessairement ses performances et mettent en avant la recherche d’un équilibre entre une modélisation trop complexe qui est inapplicable et une modélisation simplifiée à l’excès dans le but de s’adapter aux données, l’empêchant de représenter la complexité du bassin versant. Gupta *et al.* (2008) insistent par ailleurs sur la nécessaire adéquation de la complexité du modèle à la quantité d’information contenue dans les données disponibles.

III.3. Critères de performance

En 1982, Linsley déplorait l’accroissement pléthorique du nombre de modèles qui a pour effet de limiter les possibilités de comparaison des résultats des expériences de modélisation. On peut malheureusement faire un constat similaire au sujet des critères de performance : les hydrologues choisissent très souvent des critères qui leur sont propres (ou raffinent des critères utilisés largement par la communauté), ce qui entraîne la même conséquence. Dans leur étude, Dawson *et al.* (2007) proposent un ensemble varié de critères pour l’évaluation de modèles hydrologiques. Parmi ces critères, les critères d’erreur quadratique. Ils se basent sur une fonction suggérée par la régression linéaire, très connue dans le domaine des statistiques, somme des erreurs quadratiques du modèle définie par l’équation I.1 (*Dawdy et O’Donnell, 1965*) :

$$F^2 = \sum_{i=1}^n (Q_{obs,i} - Q_{cal,i})^2 \dots\dots\dots (III-1)$$

F² est analogue à la variance résiduelle d’une régression linéaire. Il est cependant difficile en utilisant F², de pouvoir comparer les performances du modèle d’une période à l’autre ou d’un bassin à l’autre, la valeur de cette fonction étant généralement d’autant plus élevée que le niveau des débits est fort (les erreurs absolues étant en moyenne plus élevées). Nash et Sutcliffe (1970) proposent donc de normaliser la fonction en effectuant une comparaison de F² avec ce même critère appliqué à un modèle élémentaire. Ils ont pris comme modèle Q = cste et ont calé ce modèle élémentaire, ce qui donne cste = Q̄_{obs}, la moyenne des débits. Pour ce modèle élémentaire, utilisé comme référence, la fonction F² devient F₀² définie comme suite (*Perrin, 2000*) :

$$F_0^2 = \sum_{i=1}^n (Q_{obs,i} - \bar{Q}_{obs})^2 \dots\dots\dots (III-2)$$

Ce qui correspond à la variance des débits observés. Nash et Sutcliffe (1970) construisent ainsi un critère d’efficacité R défini comme la proportion de la variance initiale des débits expliquée:-

$$R = 1 - \frac{F^2}{F_0^2} = 1 - \frac{\sum_{i=1}^n (Q_{obs,i} - Q_{cal,i})^2}{\sum_{i=1}^n (Q_{obs,i} - \bar{Q}_{obs})^2} \dots\dots\dots(III-3)$$

Ce critère adimensionnel (que nous appellerons critère de Nash-Sutcliffe ou critère de Nash), variant dans l'intervalle $]-\infty, 1]$ à l'avantage d'être d'interprétation facile. Il estime l'amélioration d'ajustement que l'on obtient en utilisant le modèle pour simuler les débits par rapport à un modèle *zéro* (modèle de référence) qui donnerait sur toute la période considérée un débit constant égal au débit moyen. Il vaut 1 (ou 100%) si le modèle est parfait (l'erreur du modèle est nulle). Une valeur de R inférieure à zéro signifie que le modèle n'explique pas mieux le comportement du bassin que le modèle de débit constant. Ce critère d'évaluation des modèles est très largement utilisé en hydrologie.

Cependant, est-il nécessaire d'employer une multitude de critères ? En effet, afin de se faire l'idée la plus juste de ses capacités prédictives. Cela suppose d'examiner le modèle sous des angles différents, c'est-à-dire par des critères de performance complémentaires. Cependant, on peut multiplier les critères en raffinant indéfiniment les critères. Outre, le fait que cette démarche diminue les capacités de communication entre hydrologues, il est intéressant de savoir si cela apporte réellement de l'information nouvelle : si plusieurs critères sont employés, il est nécessaire d'évaluer les corrélations entre ces critères (*Weglarczyk, 1998*).

III.4. Objet et enjeux de la modélisation pluie-débit

La modélisation pluie-débit peut réussir à répondre à de nombreuses problématiques centrées sur l'eau. Michel (1989) ; Refsgaard et Abbott (1996) répertorient l'essentiel de ces problématiques à savoir :

- *Simulation de débits*, pour le comblement de lacunes dans des séries de données, la reconstitution de débits historiques (les données de pluie étant souvent disponibles sur des périodes beaucoup plus longues que les débits) (*Sefton et Howarth, 1998*) ;
- *Prédétermination des débits de crue ou d'étiage* : on désire savoir avec quelle fréquence des débits de crue (supérieurs à un seuil de risque) ou des faibles débits (en deçà d'un débit réservé) risquent de se produire, et sur quelle durée. Cette connaissance peut permettre le dimensionnement d'ouvrages et de réservoirs ou d'aménagements dans le lit (mineur à majeur) du cours d'eau ;
- *Prévision des crues et des étiages* (*Moore, 1999*) : il s'agit d'évaluer par avance (avec un délai de quelques heures à quelques jours), connaissant l'état du bassin, les débits de crues susceptibles de présenter des risques (inondation) ou les débits d'étiages pouvant demander de mettre en place une gestion particulière de la ressource (par des barrages-réservoirs par exemple) pour assurer l'approvisionnement en eau ou la préservation de la vie halieutique ;

- **Influence d'aménagements sur l'hydrologie** : on désire pouvoir prédire les changements de la réponse du bassin suite à des modifications des caractéristiques du bassin d'origine humaine ou à des changements environnementaux.

Ces problématiques font ressortir deux aspects importants, celui de l'évaluation du risque et celui de la gestion de la ressource. La pertinence des réponses que l'on peut leur apporter est conditionnée par celle du modèle dans sa représentation du bassin relativement aux objectifs fixés. De nombreux auteurs ont étudié ces problèmes en utilisant des modèles pluie-débit. A titre d'exemple, nous pouvons citer les travaux de Cameron *et al.* (1999) et Uhlenbrook *et al.* (1999) sur la prédétermination des crues, ceux de Xu et Vandewiele (1995) et Yang *et al.* (1995) respectivement sur le dimensionnement et la gestion de réservoirs, ou ceux de Da Ros et Borga (1997) et Yang et Michel (2000) sur la prévision des débits.

III.5. Calage, validation

Différentes procédures de la modélisation de la relation pluie-débit ont été proposées dans la littérature. Il est cependant possible de trouver une structure commune à l'ensemble des procédures proposées. Elle est composée des étapes suivantes (Fig. III.4) :

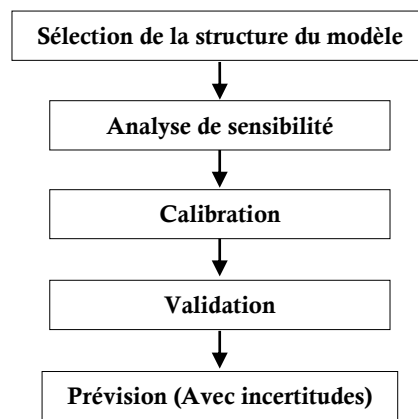


Fig. III.4. Procédure (standard) simplifiée pour la modélisation de la relation pluie-débit (Wagener *et al.*, 2004).

Tous les modèles existants nécessitent un calage. Ce calage se fait en modifiant les paramètres, à partir d'une estimation initiale, pour que le modèle ait ses sorties les plus proches possibles de celles du bassin versant à modéliser. Au sens strict du terme, c'est l'opération qui consiste à trouver les valeurs des paramètres du modèle qui minimisent l'erreur de modélisation. Cette opération peut se faire manuellement ou automatiquement (Madsen, 2000 ; Cudennec, 2000).

Une fois le modèle hydrologique implémenté sur un bassin versant donné, se pose ensuite la question de savoir quelle est la capacité du modèle à simuler le présent ou le futur ? C'est l'étape de validation qui doit répondre à cette question. Cette étape consiste à vérifier la reproductibilité des résultats par le modèle et la représentativité des paramètres calés (Gnouma, 2006).

Plusieurs schémas de calage et de validation ont été ainsi proposés par la communauté scientifique telle que le schéma de Klemeš (1983). Deux approches sont très utilisées par les hydrologues :

- Une approche classique de validation qui consiste à choisir aléatoirement une partie des données pour le calage des paramètres du modèle et une seconde partie pour la validation sans changement des valeurs des paramètres. C'est l'étape de *validation interne* qui a pour objectif d'analyser la sensibilité du modèle aux valeurs des paramètres qui le constituent ;
- Une approche basée sur la validation multicritère et multi-échelle appliquée à l'ensemble du bassin. C'est l'étape de *validation externe*. Elle consiste à confronter les résultats du modèle à une réalité. C'est-à-dire concrètement à des données (hydrogrammes mesurés sur des stations intermédiaires, niveaux piézométriques, etc.) qui n'ont pas été utilisées pour le construire. Il importe de bien veiller à ce que la validation externe d'un modèle ne soit pas conduite sur des données qui ont servi à l'établir.

III.6. Classification des modèles pluie-débit

L'approche de modélisation pluie-débit sont diverses, il existe une pléthore de modèles (Maréchal, 2004). Beven (2001) a mentionné sa tentative d'établir une liste exhaustive de modèles pluie-débit près de 25 ans et son abandon de la tâche quand il a atteint 100 modèles. Cette abondance est une conséquence légitime suite aux différentes approches possibles sur la manière de représenter la réponse d'un bassin versant à des événements pluviométriques.

Il serait hasardeux de vouloir reprendre ici une classification des types de modèles hydrologiques tant il y a presque autant de classifications que d'auteurs qui ont abordé le sujet. Nous allons néanmoins rappeler quelques termes récurrents pour justifier notre démarche. Si l'on devait n'utiliser qu'un seul terme pour caractériser un modèle hydrologique, il serait :

III.6.1. Déterministe ou stochastique

Un modèle déterministe (*Deterministic model*) s'appuie sur la description du bassin versant et prend en compte une ou plusieurs de ses caractéristiques physiques (pente, superficie, type de sol, occupation du sol, etc.). Les modèles déterministes ont pour application la prévision hydrologique pour la gestion des ouvrages hydrauliques ou pour la prévision des crues, la délimitation des zones inondables, la gestion du territoire, la quantification des impacts potentiels des changements climatiques (Fouchier, 2010).

Un modèle stochastique (*stochastic model*) sert à établir des critères de conception d'ouvrages (par exemple, le dimensionnement des évacuateurs de crue) ou à vérifier les performances d'ouvrages existants. Jensen (1993) considère qu'une approche stochastique est un moyen

rationnel de traiter la caractérisation spatiale de la variabilité, et d'établir un lien entre les incertitudes des paramètres et celles des prédictions. Les difficultés rencontrées avec ce type de modèle sont d'une part les incertitudes liées aux extrapolations quand les séries d'observations utilisées sont courtes, et d'autre part l'impossibilité d'étudier, par le biais de ces modèles, les conséquences des modifications des bassins versants (*Michel, 1989*).

III.6.2. Global, distribué ou semi-distribué

Le modèle global (*Lumped model*) (Fig. III.5) décrit le comportement du bassin dans son ensemble à partir des observations faites sur le terrain. Cette approche grandeur nature permet une certaine souplesse en ce sens qu'elle permet de tester différentes hypothèses à partir des données obtenues directement à l'échelle du bassin. On ne cherche pas à aller dans le détail des lois qui régissent le système, mais plutôt à trouver des relations entre les différentes variables du système (*Dooge, 1973 ; Sivapalan et al., 2003*).

L'un des plus graves dangers qui menace cette approche, c'est la tentation de la *théorie unitaire*, du modèle englobant ayant réponse à tout, capable de tout prévoir. De Rosnay (1975) prône donc l'approche systémique opérationnelle comme une des voies permettant d'éviter les dangereux écueils du réductionnisme paralysant et du systémisme englobant. Parmi les modèles globaux qui existent dans la littérature on peut citer : le modèle GR (*Edijatno et Michel, 1989*), Les modèles GR4J (*Perrin, 2000*), TANK (*Sugawara, 1979*), MODGLO (*Servat, 1986*), GRHUM (*Loumagne et al., 1996*).

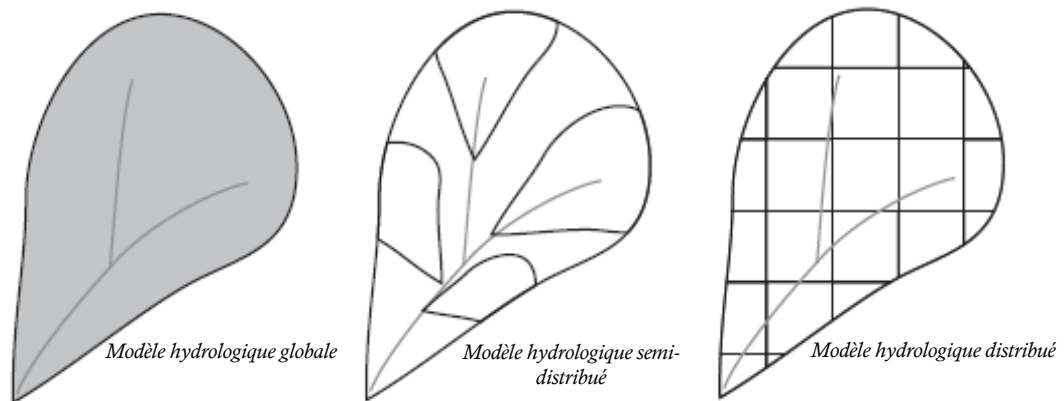


Fig. III.5. Représentations schématiques des modèles hydrologiques globales, semi-distribué et distribué (*Lerat, 2009*).

Le modèle distribuée (*Distributed model*) (Fig. III.5) dérive probablement du second précepte de Descartes énoncé dans son célèbre discours de la méthode. Il s'agit « ...de diviser chacune des difficultés que j'examinais en autant de parcelles qu'il se pourrait et qu'il serait requis pour mieux les résoudre » (*Descartes, 1637*). Les données sont donc recueillies et traitées sur des unités parcellaires (ou mailles) à petites échelles. L'avantage d'une telle approche est qu'elle permet une description

détaillée à petite échelle. S'ensuit une agrégation des informations obtenues sur chaque unité pour avoir la réponse de l'ensemble du bassin. L'approche analytique cherche à ramener le bassin à ses éléments constitutifs les plus simples afin de les étudier en détail et de comprendre les types d'interactions qui existent entre eux. Puis, en modifiant *une variable à la fois*, d'en déduire des lois générales permettant de prédire les propriétés dans des conditions très différentes (*De Rosnay, 1975*). Néanmoins, ce type de modèle pose de nombreux problèmes. A titre d'exemple, nous pouvons citer le problème de surparamétrisation. Le nombre de paramètres possible est en effet croissant avec l'augmentation du nombre de mailles ou d'unités de modèle (*Beven, 1989*).

On peut citer en exemple des modèles tels SHE d'Abbott *et al.* (1986), HYDROTEL de Fortin *et al.* (2001), IHDM de Rogers *et al.* (1985), SLURP Model de Kite (1995), TOPKAPI Model de Todini (1995), HMS de Yu (1996), DHSVM de Wigmosta *et al.* (1994) et SHETRAN de Ewen *et al.* (2000).

Le modèle semi-distribué (*semi-distributed model*) (Fig. III.5) est intermédiaire entre les deux précédents. Le bassin est généralement divisé en deux, trois ou quatre sous-bassins analysés individuellement à l'aide d'un modèle global. Les caractéristiques hydrologiques des sous-bassins sont relativement plus homogènes que le bassin pris dans son entier.

III.6.3. Empirique, conceptuel ou théorique

Lorsqu'on parle d'approche empirique (*empirical, lumped model*), il s'agit d'établir des relations entre la grandeur à modéliser et les facteurs qui la déterminent mais on n'est jamais sûr que les facteurs retenus expliquent à eux seuls la grandeur à modéliser. En plus, les facteurs qui déterminent la grandeur à modéliser ne sont pas toujours mesurés ni même souvent tous mesurables : on reste donc dans un cadre statistique (*Dreyfus et al., 2002*). C'est pourquoi on parle souvent de modèles probabilistes. En hydrologie, ces modèles sont pour la plupart globaux. Le modèle empirique ou *boîte noire* est une description paramétrée du processus dont tous les paramètres doivent être déterminés à partir des mesures effectuées. Bâtie essentiellement à partir de l'analyse des données, cette approche fonctionne très bien en interpolation mais est moins appropriée en extrapolation. L'absence d'idée préconçue dans leur construction leur permet une évolution plus rapide vers des réponses efficaces à des problématiques données. Leur développement reste cependant très dépendant de la disponibilité des données. Des exemples de ce type de modèles sont les réseaux de neurones artificiels (Artificial Neural Networks) (*Hsu et al., 1995*), les modèles GR (*Edijando et Michel, 1989 ; Loumagne, 1988*) et le modèle SCS. Ces modèles considèrent généralement le bassin versant comme une entité unique (*Bessiere, 2008*).

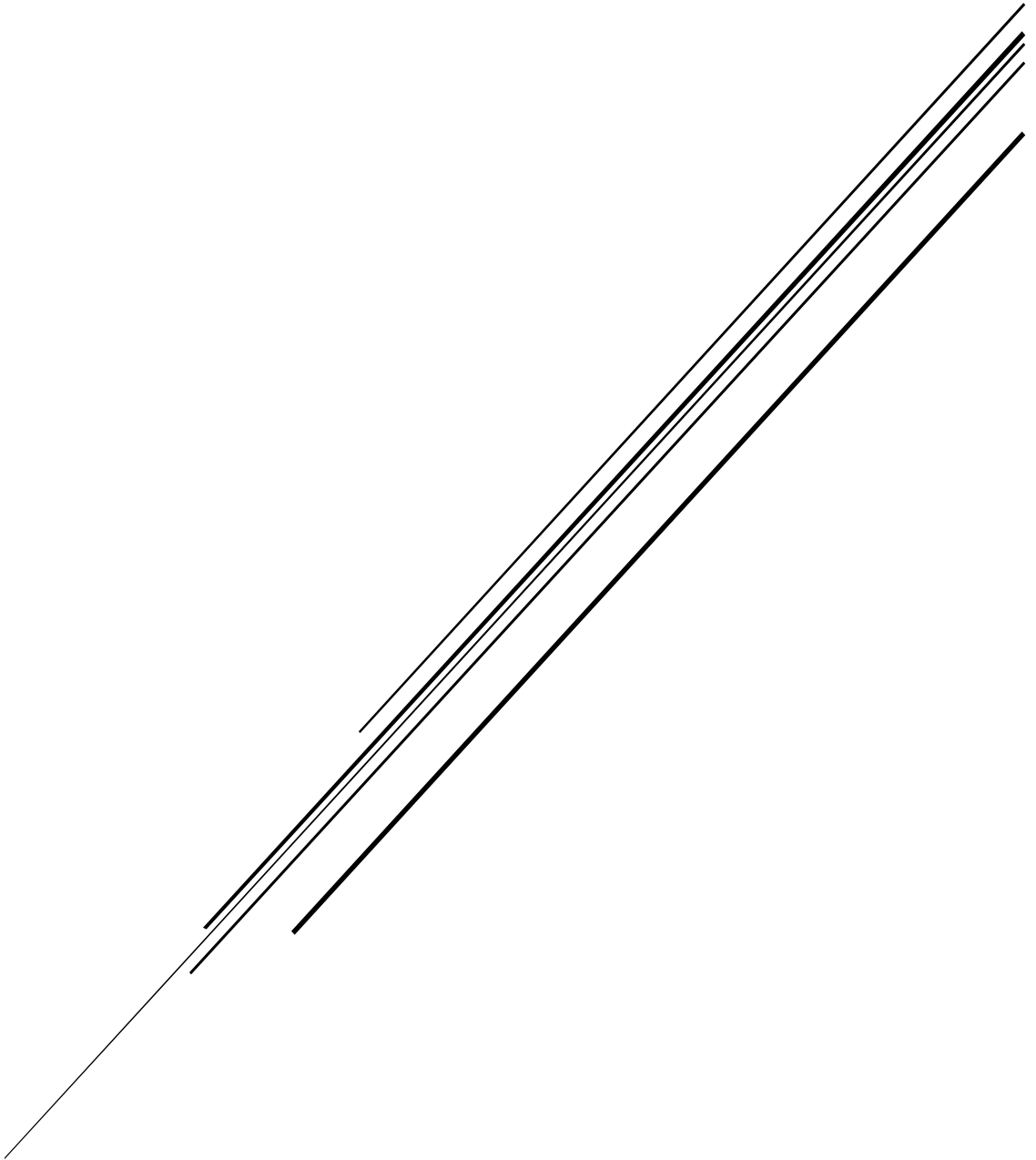
Les modèles conceptuels (*conceptual, grey box*) ne se fondent pas sur des lois hydrologiques en quelque sorte. C'est peut-être la raison pour laquelle ils sont appelés conceptuels (Yang, 1993). Le modèle conceptuel tente d'intégrer la complexité des processus en essayant de reproduire plus simplement le concept physique du comportement du système. Ambroise (1998) considère le bassin versant comme un assemblage de réservoirs d'humidité, interconnectés (on parle aussi de modèle à réservoirs) et qui censés représenter plusieurs niveaux de stockages, suivant une dimension verticale. Ce type de modèle garde un sens physique même s'il ne représente pas une description exacte de l'ensemble des processus en jeu (Chkir, 1994). Le GR4J est un exemple de modèle conceptuel, pour des applications en ingénierie (dimensionnement d'ouvrage, prévision des crues et des étiages, gestion de réservoirs) sur des bassins jaugés et non jaugés.

Les modèles théoriques sont régis par des équations mathématiques paramétrées bien connues. Encore appelés *modèles de connaissance* ou *modèles physiques* (*physically-based, white box*), ils sont fondés sur les principes mécanistes de conservation de la masse, du moment et de l'énergie. Concrètement, les équations qui génèrent ces modèles sont issues des lois de la mécanique des fluides ou de la physique des écoulements d'eau dans les sols (loi de Richards en milieu non saturé, loi de Darcy en milieu saturé et loi de Barré Saint-Venant pour le ruissellement de surface). Ils ont été développés dans l'espoir que la représentation physique de leurs processus était suffisamment réelle pour que leurs paramètres soient directement reliés aux caractéristiques physiques du bassin versant sans processus de calibration (Todini, 1988). Par exemple, ces paramètres sont représentatifs de l'humidité du sol, de la conductivité hydraulique dans la zone non saturée, etc.

Les modèles à base physique utilisent une distribution spatiale basée sur des unités élémentaires : une maille carrée régulière ou irrégulière, des triangles irréguliers TINS (Triangular Irregular Networks) (Ivanov et al., 2004), des tubes de courant (Grayson et al., 1992), des versants ou d'autres types d'unités hydrologiques. Cette distribution spatiale permet de représenter la variabilité spatiale des facteurs géographiques et météorologiques qui déterminent la réponse hydrologique. Le modèle SHE (Abbott et al., 1986) et le modèle TOPKAPI (Liu et Todini, 2005) sont les exemples les plus connus. Cependant, les modèles à base physique souffrent d'une extrême demande de données, des problèmes d'échelles et de surparamétrisation.

Pour finir, il ne devrait pas y avoir de *conflicts* entre ces différents modèles qui ont été développés à des fins différentes (Bergström, 1991). Ainsi, les modèles théoriques sont plutôt utilisés pour une étude fine des processus hydrologiques tandis que les modèles conceptuels correspondent à une approche quasi statistique, où les paramètres correspondent à une moyenne sur l'ensemble du bassin (Miossec, 2004).

REFERENCES BIBLIOGRAPHIQUES



REFERENCES BIBLIOGRAPHIQUES

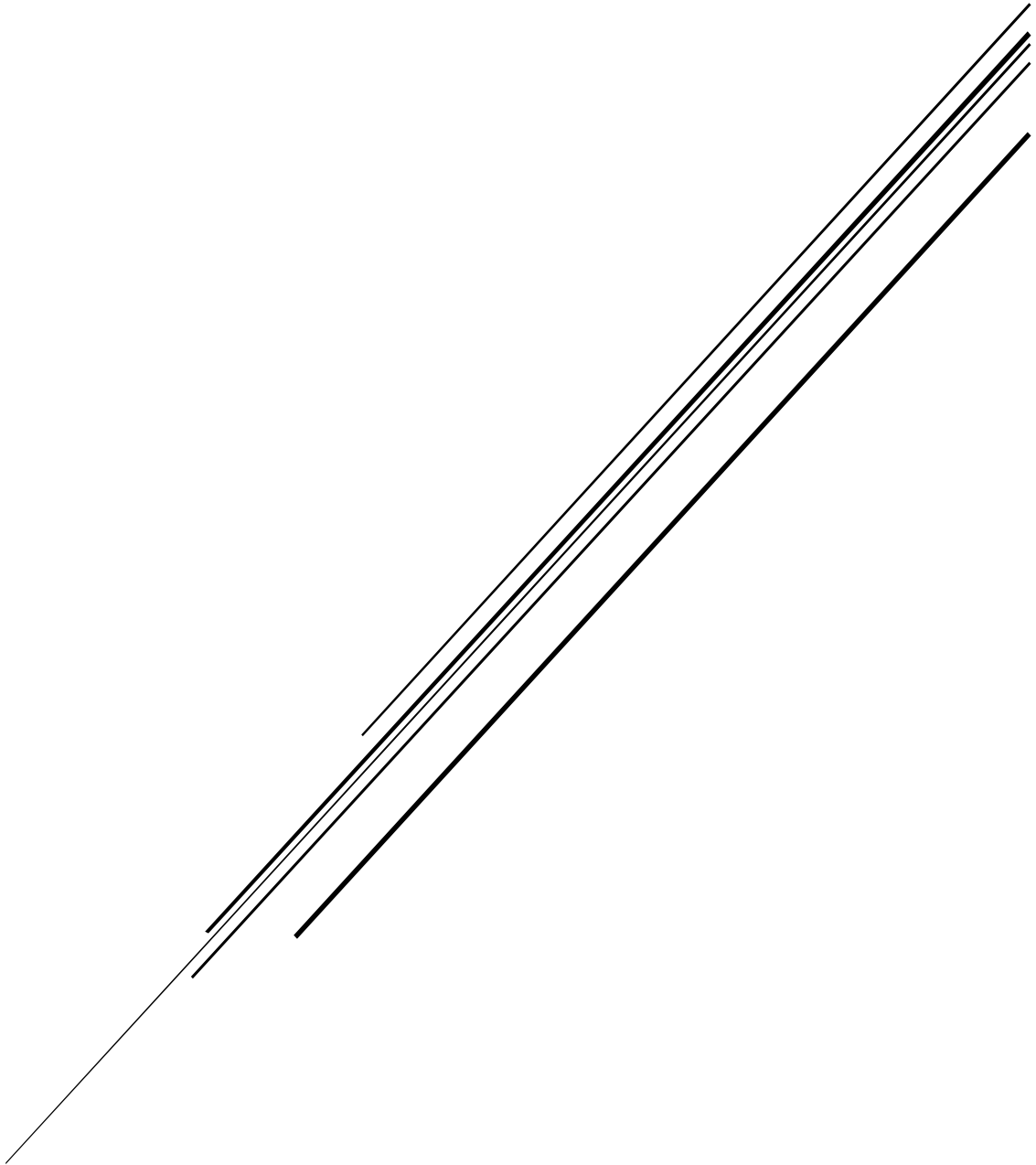
- Abbott M.B., Bathurst J.C., Cunge J.A., O'Connell P.E., Rasmussen J.**, 1986. An Introduction to the European Hydrological System - Système Hydrologique Européen, "SHE", 2: Structure of a physically-based, distributed modeling system. *Journal of Hydrology*, 87, 61-77.
- Ambroise B.**, 1998. La dynamique du cycle de l'eau dans un bassin versant - Processus, Facteurs, Modèles, Ed., HGA, Bucarest.
- Ambroise B.**, 1999. Genèse des débits – 2. Modélisation. *Revue des Sciences de l'eau*, 12(1), 123-153.
- Andersson L.**, 1992. Improvements of ruff models – What's way to go? *Nordic Hydrology*, 23(5), 315-332.
- Anonyme**. 2015a. Hydrologie Statistique 1. Cours Aménagement et Ouvrages hydrotechniques. Master 1.
- Anonyme**. 2015b. Hydrologie Statistique 2. Cours Aménagement et Ouvrages hydrotechniques. Master 1.
- Anonyme**. Prévion des crues et risques d'inondation. Cours. Disponible sur : <http://agronomies.blogspot.com/>
- Anonyme**. Série des travaux dirigés en hydrologie (exercices et corrigés). USTHB – FGC –1MIE-1112.
- Banks J.**, 1998. Handbook of simulation: principles, methodology, advances, applications and practice. New York: Wiley Interscience
- Belarbi H., Matari A., Habi M.** 2012. Etude des séries temporelles. Applications aux données hydroclimatologiques. Edition Universitaires Européennes. ISBN-13.978-3-8417-9316-4
- Bergström S.**, 1991. Principles and Confidence in Hydrological Modelling. *Nordic hydrology*, 22, 123-136.
- Bernier J., Veron R.** 1964. Sur quelques difficultés rencontrées dans l'estimation d'un débit de crue de probabilité donnée. *Revue de statistique appliquée*. 12(1) :25-48. Disponible sur : http://www.numdam.org/article/RSA_1964__12_1_25_0.pdf
- Bessiere H.**, 2008. Assimilation de données variationnelle pour la modélisation hydrologique distribuée des crues à cinétique rapide. Thèse de Doctorat. Université de Toulouse. Institut National Polytechnique de Toulouse.
- Beven K. J.**, 1989. Changing ideas in hydrology — The case of physically-based models. *Journal of Hydrology*, 105, 157-172.
- Beven K.J.**, 2001. In *Rainfall-Runoff modeling the Primer*. John Wileys & Sons. Ltd.
- Bois Ph., Obled Ch., Zin I.** 2007. Introduction au traitement de données en hydrologie. L'Édition du Millénaire. 7ème édition revue et complétée. p 265.
- Cameron D.S., Beven K.J., Tawn J., Blazkova S., Naden P.**, 1999. Flood frequency estimation by continuous simulation for a gauged upland catchment (with uncertainty). *Journal of Hydrology*, 219, 169-187.
- Chkir N.**, 1994. Mise au point d'un modèle hydrologique conceptuel intégrant l'état hydrique du sol dans la modélisation pluie-débit. Thèse de Doctorat. Ecole Nationale des Ponts et Chaussées.
- Cudennec C.**, 2000. Description mathématique de l'organisation du réseau hydrographique et modélisation hydrologique. Thèse de Doctorat. Ecole Nationale Supérieur Agronomique. Rennes.
- Da Ros D., Borga M.**, 1997. Adaptive use of a conceptual model for real time forecasting. *Nordic Hydrology*, 28(3), 169-188
- Dawdy D.R., O'Donnell T.**, 1965. Mathematical models of catchment behavior. *American Society of Civil Engineers Proceedings*, 4410 (HY4), 123-137
- Dawson C., Abrahart R., See L.**, 2007. Hydro Test: A web-based toolbox of evaluation metrics for the standardised assessment of hydrological forecasts. *Environmental Modelling and Software*, 22(7), 1034-1052.
- De Rosnay J.**, 1975. Le microscope. Vers une vision globale. Seuil. (Prix de l'Académie des sciences morales et politiques). Paris
- Descartes R.**, 1637. Discours de la méthode. Ed. Librairie Philosophique. Paris.

- Dooge J.C.I.**, 1973. Linear theory of hydrology systems. Technical Bulletin.
- Dreyfus G., Martinez J.M., Samuelides M., Gordon M.B., Badran F., Thiria S., Hérault L.**, 2002. Réseaux de neurones: Méthodologie et applications. Ed. Eyrolles
- Duband D.**, 1982. Hydrologie statistique approfondie. Cours. Institut National Polytechnique de Grenoble. p 276.
- Edijatno C., Michel C.**, 1989. Un modèle pluie-débit journalier à trois paramètres. La Houille Blanche, 2, 113-121.
- Ewen J., Parkin G., O'Connell P.E.**, 2000. SHETRAN: Distributed River Basin Flow and Transport Modeling System. Journal of Hydrologic Engineering, 5(3), 250-258
- Fenicia F., Savenije H., Matgen P., Pfister L.**, 2008. Understanding catchment behavior through stepwise model concept improvement. Water Resources Research, 44(1): W01402. Doi : 10. 1029/2006WR005563
- Fortin J.P., Turcotte R., Massicotte S., Moussa R., Fitzback J., Villeneuve J.**, 2001. Distributed Watershed Model Compatible with Remote Sensing and GIS Data. II : Application to Chaudière Watershed. Journal of Hydrologic Engineering, 6(2), 100-108.
- Fouchier C.**, 2010. Développement d'une méthodologie pour la connaissance régionale des crues. Thèse de Doctorat. Université des Sciences et Technique du Languedoc. Montpellier II. CEMAGREF
- Gaume E.**, 2002. Eléments d'analyse sur les crues éclairs. Thèse de Doctorat. ENGREF. INRS.
- Gérard L.**, 2010. Sensibilité des performances d'un modèle de prévision des crues au critère de calage. Rapport de stage. CEMAGREF.
- Gnouma R.**, 2006. Aide à la calibration d'un modèle hydrologique distribué à moyen d'une analyse des processus hydrologiques : application au bassin versant de l'Yzeron. Thèse de Doctorat. Unité de recherche en hydrologie-hydraulique, CEMAGREF
- Goldfarb B., Pardoux C.** 2013. Introduction à la méthode statistique. Statistique et probabilité (cours et exercices corrigés). Edition Dunod (7e édition).p 385.
- Grayson R.B., Moore I.D., McMahon T.A.**, 1992. Physically-based hydrologic modeling. 2. Is the concept realistic? Water Resources Research, 26(10), 2659-2666
- Gupta H.V., Wagener T., Liu Y.**, 2008. Reconciling theory with observations: Elements of a diagnostic approach to model evaluation. Hydrological Processes, 22(18), 3802-3813.
- Gupta V.K., Sorooshian S.**, 1985. The relationship between data and the precision of parameter estimates of hydrologic models. Journal of Hydrology, 81, 57-77.
- Hsu K.L., Gupta H.V., Sorooshian S.**, 1995. Artificial neural network modeling of the rainfall-runoff process. Water Resources Research, 31(10), 2517-2530
- Ivanov V.Y., Vvoni E.R., Bras R.L., Entekhabi D.**, 2004. Catchment hydrologic response with a fully distributed triangulated irregular network model. Water Resources Research, 40(W11102), 1-23. doi :10.1029/2004WR003218
- Jakeman A.J., Hornberger G.M.**, 1993. How much complexity is warranted in a rainfall runoff model? Water Resources Research, 29(8), 2637-2649
- Jensen R.C.**, 1993. Effect of Forestry on suspended solids and bedload yields in the Balquhiddy catchment. Journal of Hydrology, 145, 403-417.
- Kauark-Leite L.A., Nascimento N.O.**, 1993. Développement, utilisation et incertitudes des modèles conceptuels en hydrologie. Séminaire Eaux - Environnement n°3. Modélisation du comportement des polluants dans les hydrosystèmes. Paris, 191-219.
- Kite G.**, 1995. The SLURP model, In: Computer Models of Watershed Hydrology. Ed. Water Resources Publications. Highlands Ranch. Colorado
- Klemeš V.**, 1983. Conceptualization and scale in hydrology. Journal of hydrology, 65, 1-23

- Laborde J.P.**, 2009. Eléments d'hydrologie de surface. Edition Ecole Polytechnique de l'Université de Nice - Sophia Antipolis. p202.
- Le Lay M.**, 2006. Modélisation hydrologique dans un contexte de variabilité hydro-climatique. Une approche comparative pour l'étude du cycle hydrologique à méso-échelle au Benin. Thèse de Doctorat. Institut National Polytechnique de Grenoble.
- Lerat J.**, 2009. Quels apports hydrologiques pour les modèles hydrauliques ? Vers un modèle intégré de simulation des crues. Thèse de Doctorat. Université Pierre et Marie Curie. Paris VI.
- Lindström G., Johansson B., Persson M., Gardelin M., Bergström S.**, 1997. Development and test of the distributed HBV-96 hydrological model. *Journal of Hydrology*, 201, 272-288.
- Linsley R.K.**, 1982. Rainfall-runoff models-an overview, In: Rainfall-runoff relationship. V. P. Singh, Ed. Water Resources Publications. Littleton, CO
- Liu Z., Todini E.**, 2005. Assessing the TOPKAPI non-linear reservoir cascade approximation by means of a characteristics lines solution. *Hydrological Processes*, 19(10), 1983-2006
- Loumagne C.**, 1988. Prise en compte d'un indice de l'état hydrique du sol dans la modélisation pluie-débit. Thèse de Doctorat. Université Paris Sud-Centre d'Orsay. CEMAGREF Antony
- Loumagne C., Chkir N., Normand M., Ottlé C., Vidal-Madjar D.**, 1996. Introduction of the soil-vegetation-atmosphere continuum in a conceptual rainfall-runoff model. *Journal of Hydrological Sciences*. 41(6), 889-902
- Madsen H.**, 2000. Automatic calibration of a conceptual rainfall-runoff model using multiple objectives. *Journal of Hydrology*, 235, 276-288
- Maréchal D.**, 2004. A soil-based approach to rainfall-runoff modelling in ungauged catchments for England and Wales. Thèse de Doctorat. Cranfield University
- Michel C.**, 1989. Hydrologie appliquée aux petits bassins versants ruraux. Ed. CEMAGREF, Antony
- Miossec M.P.**, 2004. Apport des Multi-Modèles pour la modélisation sur des bassins versants non jaugés. Mémoire de DEA. Université Pierre et Marie Curie. Université Paris-Sud. École des Mines de Paris. ENGREF
- Moore R.J.**, 1999. Real-time flood forecasting systems: Perspectives and prospects. In *Floods and landslides: integrated Risk Assessment*. Ed. Floods and landslides: Integrated risk assessment. Springer-Verlag. Berlin
- Nash J.E., Sutcliffe J.V.**, 1970. River flow forecasting through conceptual models. Part I-A discussion of principles. *Journal of Hydrology*, 27(3), 282-290.
- Perrin C.**, 2000. Vers une amélioration d'un modèle global pluie-débit au travers d'une approche comparative. Thèse de Doctorat. Institut National Polytechnique de Grenoble
- Refsgaard J.C., Abbott M.B.**, 1996. The role of distributed hydrological modelling in water resources management - Chapter 1. In: *Distributed Hydrological modeling*. M. B. Abbott et J. C. Refsgaard (Ed.), Kluwer Academic Publishers, Netherlands
- Roche M.** 1963. Hydrologie de surface. Edition Gauthier-Villars Editeur- Office de La Recherche Scientifique et Technique Outre-Mer (ORSTOM). p 431.
- Roche P. A.**, 1988. Les modèles. In : Valiron, F (ed). *Gestion des eaux : Automatisation, Informatisation et Télégestion*. Tome III. Presses de l'Ecole Nationale des Ponts et Chaussées.
- Rogers C.C.M., Beven K.J., Morris E.M., Anderson M.G.**, 1985. Sensitivity analysis, calibration and predictive uncertainty of the Institute of Hydrology Distributed Model. *Journal of Hydrology*, 81, 179-191
- Rosen R.**, 1991. A comprehensive inquiry into the nature, origin and fabrication of life. New York: Columbia University Press
- Sari A.A.** 2002a. Initiation à l'hydrologie de surface. Cours. Editions- Distribution Houma. Réf. : 5/080. p 222.
- Sari A.A.** 2002b. Initiation à l'hydrologie de surface. Exercices et corrigés. Editions- Distribution Houma. Réf. : 5/080. p 154.
- Sefton C.E.M., Howarth S.M.**, 1998. Relationships between dynamic response characteristics and physical descriptors of catchments in England and Wales. *Journal of Hydrology*, 211, 1-16.

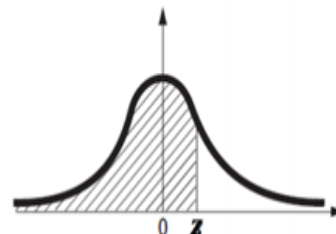
- Servat E.**, 1986. Présentation de Trois Modèles Globaux Conceptuels Déterministes : CREC5, MODGLO et MODIBI. Rapport ORSTOM
- Sivapalan M., Bloschl G., Zhang L., Vertessy R.**, 2003. Downward approach to hydrological prediction. *Hydrological Processes*, 17, 2101-2111
- Sugawara M.**, 1979, Automatic calibration of the tank model. *Hydrological Sciences Bulletin*, 24(3), 375-388
- Thiery D.** 1989. *Eléments d'analyse statistique. Application à l'hydrologie.* 2ème Edition. Bureau de Recherches Géologiques et Minières. Services sol et sous-sol (BRGM). Département de l'Eau. p 75.
- Thom R.**, 1979. Modélisation et Scientificité. In, Delattre, P et Thellier M, (ed), Actes du colloque « Elaboration et Justification des modèles : Application en Biologie ». Tome I. Ed, Malione, Paris
- Todini E.**, 1988. Rainfall-runoff modelling - Past, present and future. *Journal of Hydrology*, 100, 341-352.
- Todini E.**, 1995. The Role of Water and the Hydrological Cycle in Global Change. Series I, 31, Oliver HR
- Touaïbia B.** 2004. Manuel pratique d'hydrologie. Dépôt légal : 567-2004. ISBN : 9947-0-0367-1. p 159.
- Touaïbia B., Laborde J-P., Brigode P.** 2018. Hydrologie méditerranéenne. Exercices corrigés. Ecole Polytechnique de Nice-Sophia Antipolis. Ecole Supérieure d'Hydraulique. Disponible sur : <https://library.ensh.dz/index.php/23-livre-du-mois/97-livre-fevrier-2018>
- Uhlenbrook S., Seibert J., Leibundgut C., Rodhe A.**, 1999. Prediction uncertainty of conceptual rainfall-runoff models caused by problems in identifying model parameters and structure. *Hydrological Sciences Journal*, 44 (5), 779-797
- Vessereau A.** 1972. La statistique. Douzième Edition. p 325.
- Wagener T., Wheater H. S., Gupta, H. V.**, 2004. *Rainfall-Runoff Modelling in Gauged and Ungauged Catchments.* Ed I. Imperial College Press. London. UK.
- Weglarczyk S.**, 1998. The interdependence and applicability of some statistical quality measures for hydrological models. *Journal of Hydrology*, 206, 98-103
- Wigmosta M.S., Vail, L., Lettenmaier, D.P.**, 1994. A distributed hydrology-vegetation model for complex terrain. *Water Resources Research*, 30(6), 1665-1679
- Xu C.Y., Vandewiele G.L.**, 1995. Parsimonious monthly rainfall-runoff models for humid basins with different input requirements. *Advances in Water Resources*, 18, 39-48
- Yang X.**, 1993. Mise au point d'une méthode d'utilisation d'un modèle pluie-débit conceptuel pour la prévision des crues en temps réel. Thèse de Doctorat. Ecole Nationale des Ponts et Chaussées
- Yang X., Michel C.**, 2000. Flood forecasting with a watershed model: a new method of parameter updating. *Hydrological Sciences Journal*. 45(4), 537-546.
- Yang X., Parent E., Michel C., Roche P.A.**, 1995. Comparison of real-time reservoir operation techniques. *Journal of Water Resources Planning and Management*, 121(5), 345-351
- Yu Z.**, 1996. Development of a physically-based distributed-parameter watershed – basin scale hydrology model! And its application to Big Darby Creek watershed. Thesis of Doctorate. State University Of Columbus. Ohio.
- Zendagui D.** 2011. Maths 4: Probabilité et Statistiques. Notes de Cours. Université de Tlemcen. p 60.
- Zin I.**, 2002. Incertitudes et ambiguïté dans la modélisation hydrologique. Discussion, développement méthodologiques et application à l'hydrologie de crue en Ardèche. Thèse de Doctorat. Institut National Polytechnique de Grenoble.

ANNEXES



Annexe I. Table loi normale centrée réduite $N(0;1)$

$$F(z) = P(Z \leq z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-\frac{x^2}{2}} dx$$



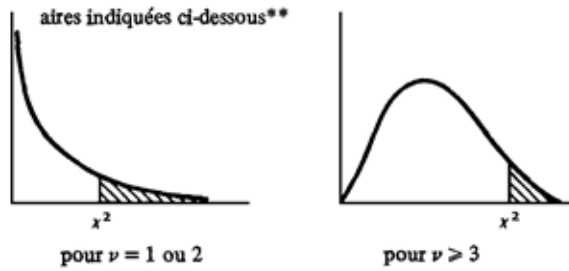
z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0	0,50000	0,50399	0,50798	0,51197	0,51595	0,51994	0,52392	0,52790	0,53188	0,53586
0,1	0,53983	0,54380	0,54776	0,55172	0,55567	0,55962	0,56356	0,56749	0,57142	0,57535
0,2	0,57926	0,58317	0,58706	0,59095	0,59483	0,59871	0,60257	0,60642	0,61026	0,61409
0,3	0,61791	0,62172	0,62552	0,62930	0,63307	0,63683	0,64058	0,64431	0,64803	0,65173
0,4	0,65542	0,65910	0,66276	0,66640	0,67003	0,67364	0,67724	0,68082	0,68439	0,68793
0,5	0,69146	0,69497	0,69847	0,70194	0,70540	0,70884	0,71226	0,71566	0,71904	0,72240
0,6	0,72575	0,72907	0,73237	0,73565	0,73891	0,74215	0,74537	0,74857	0,75175	0,75490
0,7	0,75804	0,76115	0,76424	0,76730	0,77035	0,77337	0,77637	0,77935	0,78230	0,78524
0,8	0,78814	0,79103	0,79389	0,79673	0,79955	0,80234	0,80511	0,80785	0,81057	0,81327
0,9	0,81594	0,81859	0,82121	0,82381	0,82639	0,82894	0,83147	0,83398	0,83646	0,83891
1	0,84134	0,84375	0,84614	0,84849	0,85083	0,85314	0,85543	0,85769	0,85993	0,86214
1,1	0,86433	0,86650	0,86864	0,87076	0,87286	0,87493	0,87698	0,87900	0,88100	0,88298
1,2	0,88493	0,88686	0,88877	0,89065	0,89251	0,89435	0,89617	0,89796	0,89973	0,90147
1,3	0,90320	0,90490	0,90658	0,90824	0,90988	0,91149	0,91309	0,91466	0,91621	0,91774
1,4	0,91924	0,92073	0,92220	0,92364	0,92507	0,92647	0,92785	0,92922	0,93056	0,93189
1,5	0,93319	0,93448	0,93574	0,93699	0,93822	0,93943	0,94062	0,94179	0,94295	0,94408
1,6	0,94520	0,94630	0,94738	0,94845	0,94950	0,95053	0,95154	0,95254	0,95352	0,95449
1,7	0,95543	0,95637	0,95728	0,95818	0,95907	0,95994	0,96080	0,96164	0,96246	0,96327
1,8	0,96407	0,96485	0,96562	0,96638	0,96712	0,96784	0,96856	0,96926	0,96995	0,97062
1,9	0,97128	0,97193	0,97257	0,97320	0,97381	0,97441	0,97500	0,97558	0,97615	0,97670
2	0,97725	0,97778	0,97831	0,97882	0,97932	0,97982	0,98030	0,98077	0,98124	0,98169
2,1	0,98214	0,98257	0,98300	0,98341	0,98382	0,98422	0,98461	0,98500	0,98537	0,98574
2,2	0,98610	0,98645	0,98679	0,98713	0,98745	0,98778	0,98809	0,98840	0,98870	0,98899
2,3	0,98928	0,98956	0,98983	0,99010	0,99036	0,99061	0,99086	0,99111	0,99134	0,99158
2,4	0,99180	0,99202	0,99224	0,99245	0,99266	0,99286	0,99305	0,99324	0,99343	0,99361
2,5	0,99379	0,99396	0,99413	0,99430	0,99446	0,99461	0,99477	0,99492	0,99506	0,99520
2,6	0,99534	0,99547	0,99560	0,99573	0,99585	0,99598	0,99609	0,99621	0,99632	0,99643
2,7	0,99653	0,99664	0,99674	0,99683	0,99693	0,99702	0,99711	0,99720	0,99728	0,99736
2,8	0,99744	0,99752	0,99760	0,99767	0,99774	0,99781	0,99788	0,99795	0,99801	0,99807
2,9	0,99813	0,99819	0,99825	0,99831	0,99836	0,99841	0,99846	0,99851	0,99856	0,99861
3	0,99865	0,99869	0,99874	0,99878	0,99882	0,99886	0,99889	0,99893	0,99896	0,99900
3,1	0,99903	0,99906	0,99910	0,99913	0,99916	0,99918	0,99921	0,99924	0,99926	0,99929
3,2	0,99931	0,99934	0,99936	0,99938	0,99940	0,99942	0,99944	0,99946	0,99948	0,99950
3,3	0,99952	0,99953	0,99955	0,99957	0,99958	0,99960	0,99961	0,99962	0,99964	0,99965
3,4	0,99966	0,99968	0,99969	0,99970	0,99971	0,99972	0,99973	0,99974	0,99975	0,99976
3,5	0,99977	0,99978	0,99978	0,99979	0,99980	0,99981	0,99981	0,99982	0,99983	0,99983
3,6	0,99984	0,99985	0,99985	0,99986	0,99986	0,99987	0,99987	0,99988	0,99988	0,99989
3,7	0,99989	0,99990	0,99990	0,99990	0,99991	0,99991	0,99992	0,99992	0,99992	0,99992
3,8	0,99993	0,99993	0,99993	0,99994	0,99994	0,99994	0,99994	0,99995	0,99995	0,99995
3,9	0,99995	0,99995	0,99996	0,99996	0,99996	0,99996	0,99996	0,99996	0,99997	0,99997

Annexe II. Fonction Rybkine Foster $F_{Pr}(Pr, C_s)$

Cs	0,01	0,1	1	3	5	10	20	25	30	40	50	60	70	75	80	90	95	97	99	99,9	5%-F95	S
0,00	3,72	3,09	2,33	1,88	1,64	1,28	0,84	0,67	0,52	0,25	0,00	-0,25	-0,52	-0,67	-0,84	-1,28	-1,64	-1,88	-2,33	-3,09	3,28	0,00
0,10	3,94	3,23	2,40	1,92	1,67	1,29	0,84	0,66	0,51	0,24	-0,02	-0,27	-0,53	-0,68	-0,85	-1,27	-1,61	-1,84	-2,25	-2,95	3,28	0,03
0,20	4,16	3,33	2,47	1,96	1,70	1,30	0,83	0,65	0,50	0,22	-0,03	-0,28	-0,55	-0,69	-0,85	-1,26	-1,58	-1,79	-2,18	-2,81	3,28	0,06
0,30	4,38	3,52	2,54	2,00	1,72	1,31	0,82	0,64	0,48	0,20	-0,05	-0,30	-0,56	-0,70	-0,85	-1,24	-1,55	-1,75	-2,10	-2,67	3,27	0,08
0,40	4,61	3,66	2,61	2,04	1,75	1,32	0,82	0,63	0,47	0,19	-0,07	-0,31	-0,57	-0,71	-0,85	-1,23	-1,52	-1,70	-2,03	-2,54	3,27	0,11
0,50	4,83	3,81	2,68	2,08	1,77	1,33	0,81	0,62	0,46	0,17	-0,08	-0,33	-0,58	-0,71	-0,85	-1,22	-1,49	-1,66	-1,96	-2,40	3,26	0,14
0,60	5,05	3,96	2,75	2,12	1,80	1,33	0,80	0,61	0,44	0,16	-0,10	-0,34	-0,59	-0,72	-0,85	-1,20	-1,45	-1,61	-1,88	-2,27	3,25	0,17
0,70	5,28	4,10	2,82	2,15	1,82	1,34	0,79	0,59	0,43	0,14	-0,12	-0,36	-0,60	-0,72	-0,85	-1,18	-1,42	-1,57	-1,81	-2,14	3,24	0,20
0,80	5,50	4,24	2,89	2,18	1,84	1,34	0,78	0,58	0,41	0,12	-0,13	-0,37	-0,60	-0,73	-0,85	-1,17	-1,38	-1,52	-1,74	-2,02	3,22	0,22
0,90	5,73	4,32	2,96	2,22	1,86	1,34	0,77	0,57	0,40	0,11	-0,15	-0,38	-0,61	-0,73	-0,85	-1,15	-1,35	-1,47	-1,66	-1,90	3,21	0,25
1,00	5,96	4,53	3,02	2,25	1,88	1,34	0,76	0,55	0,38	0,09	-0,15	-0,39	-0,62	-0,73	-0,85	-1,13	-1,32	-1,42	-1,59	-1,79	3,20	0,28
1,10	6,18	4,67	3,09	2,28	1,89	1,34	0,74	0,54	0,36	0,07	-0,18	-0,41	-0,62	-0,74	-0,85	-1,10	-1,28	-1,38	-1,52	-1,68	3,17	0,31
1,20	6,41	4,81	3,15	2,31	1,91	1,34	0,73	0,54	0,35	0,05	-0,19	-0,42	-0,63	-0,74	-0,85	-1,08	-1,24	-1,33	-1,45	-1,58	3,16	0,31
1,30	6,64	4,95	3,21	2,34	1,92	1,34	0,72	0,49	0,33	0,04	-0,21	-0,43	-0,63	-0,74	-0,84	-1,06	-1,20	-1,28	-1,38	-1,48	3,14	0,37
1,40	6,87	5,09	3,27	2,37	1,94	1,33	0,71	0,47	0,31	0,02	-0,22	-0,44	-0,64	-0,73	-0,83	-1,04	-1,17	-1,23	-1,32	-1,39	3,12	0,39
1,50	7,09	5,23	3,33	2,39	1,95	1,33	0,69	0,46	0,30	0,00	-0,24	-0,45	-0,64	-0,73	-0,82	-1,02	-1,13	-1,19	-1,26	-1,31	3,09	0,42
1,60	7,31	5,37	3,39	2,42	1,96	1,32	0,68	0,44	0,25	-0,02	-0,25	-0,46	-0,64	-0,73	-0,81	-0,99	-1,10	-1,14	-1,20	-1,24	3,07	0,45
1,70	7,54	5,50	3,44	2,44	1,97	1,32	0,66	0,42	0,25	-0,03	-0,27	-0,47	-0,64	-0,72	-0,80	-0,97	-1,06	-1,10	-1,14	-1,17	3,04	0,48
1,80	7,76	5,64	3,50	2,46	1,98	1,31	0,64	0,40	0,24	-0,05	-0,28	-0,48	-0,64	-0,72	-0,79	-0,94	-1,02	-1,06	-1,09	-1,11	3,01	0,51
1,90	7,98	5,77	3,55	2,49	1,99	1,30	0,63	0,39	0,22	-0,07	-0,03	-0,48	-0,64	-0,71	-0,78	-0,92	-0,98	-1,01	-1,04	-1,05	2,98	0,54
2,00	8,21	5,91	3,60	2,51	2,00	1,29	0,61	0,37	0,20	-0,05	-0,31	-0,49	-0,64	-0,70	-0,78	-0,90	-0,95	-0,97	-0,99	-1,00	2,95	0,57
2,10	--	6,04	3,65	2,53	2,01	1,27	0,59	0,35	0,18	-0,10	-0,32	-0,50	-0,64	-0,69	-0,75	-0,81	-0,91	-0,99	-0,95	-0,95	2,92	0,59
2,20	--	6,14	3,68	2,54	2,02	1,26	0,57	0,29	0,16	-0,12	-0,33	-0,50	-0,64	-0,68	-0,74	-0,84	-0,88	-0,90	-0,91	-0,91	2,89	0,62
2,30	--	6,26	3,73	2,57	2,01	1,25	0,55	0,27	0,14	-0,13	-0,34	-0,50	-0,63	-0,67	-0,72	-0,82	-0,85	-0,83	-0,87	-0,87	2,86	0,64
2,40	--	6,37	3,78	2,60	2,03	1,23	0,52	0,25	0,14	-0,14	-0,35	-0,51	-0,62	-0,66	-0,71	-0,79	-0,82	-0,80	-0,83	0,83	2,82	0,67
2,50	--	6,50	3,82	2,62	2,00	1,21	0,50	0,24	0,10	-0,16	-0,36	-0,51	-0,62	-0,66	-0,70	-0,77	-0,79	-0,77	-0,80	-0,80	2,79	0,69
2,60	--	6,54	3,86	2,63	2,00	1,19	0,48	0,22	0,04	-0,17	-0,37	-0,51	-0,61	-0,65	-0,70	-0,75	-0,76	-0,74	-0,77	-0,77	2,76	0,72
2,70	--	6,75	3,92	2,64	2,00	1,18	0,46	0,20	0,07	-0,18	-0,38	-0,51	-0,61	-0,64	-0,68	-0,72	-0,74	-0,71	-0,74	-0,74	2,74	0,74
2,80	--	6,86	3,96	2,64	2,00	1,15	0,44	0,19	0,06	-0,20	-0,39	-0,51	-0,60	-0,63	-0,67	-0,71	-0,71	-0,69	-0,71	-0,71	2,71	0,76
2,90	--	7,00	4,01	2,65	1,99	1,13	0,41	0,17	0,04	-0,21	-0,39	-0,05	-0,60	-0,62	-0,65	-0,68	-0,69	-0,67	-0,69	-0,69	2,68	0,78
3,00	--	7,10	4,05	2,66	1,97	1,11	0,39	0,15	0,03	-0,22	-0,40	-0,51	-0,59	-0,60	-0,64	-0,66	-0,67	-0,65	-0,67	-0,67	2,64	0,80
3,10	--	7,23	4,09	2,66	1,97	1,09	0,37	0,13	0,01	-0,23	-0,40	-0,51	-0,58	-0,59	-0,62	-0,64	-0,64	-0,63	-0,65	-0,65	2,62	0,81
3,20	--	7,35	4,11	2,64	1,96	1,08	0,35	0,11	-0,01	-0,25	-0,41	-0,51	-0,57	-0,59	-0,61	-0,62	-0,63	-0,61	-0,63	-0,63	2,59	0,83
3,30	--	7,44	4,15	2,60	1,95	1,06	0,33	0,09	-0,02	-0,26	-0,41	-0,51	-0,56	-0,58	-0,59	-0,60	-0,61	-0,59	-0,61	-0,61	2,56	0,85
3,40	--	7,54	4,18	2,66	1,95	1,04	0,31	0,06	-0,04	-0,27	-0,41	-0,50	-0,55	-0,57	-0,58	-0,59	-0,59	-0,57	-0,58	-0,59	2,53	0,86
3,50	--	7,64	4,21	2,66	1,93	1,03	0,29	0,05	-0,05	-0,28	-0,41	-0,50	-0,54	-0,55	-0,56	-0,57	-0,57	-0,56	-0,57	-0,57	2,50	0,87
3,60	--	7,72	4,24	2,66	1,93	1,01	0,28	0,03	-0,07	-0,28	-0,42	-0,50	-0,54	-0,54	-0,56	-0,56	-0,56	-0,54	-0,56	-0,55	2,48	0,89
3,70	--	7,86	4,26	2,66	1,91	1,01	0,26	0,03	-0,08	-0,29	-0,42	-0,49	-0,52	-0,53	-0,55	-0,54	-0,54	-0,53	-0,54	-0,54	2,45	0,90
3,80	--	7,97	4,29	2,65	1,90	1,00	0,24	0,03	-0,10	-0,30	-0,42	-0,48	-0,51	-0,52	-0,54	-0,53	-0,53	-0,53	-0,53	-0,53	2,43	0,91

Annexe III. Table de distribution du χ^2

Table* des valeurs de χ^2 en fonction du nombre ν de degrés de liberté et de l'aire relative à droite de χ^2

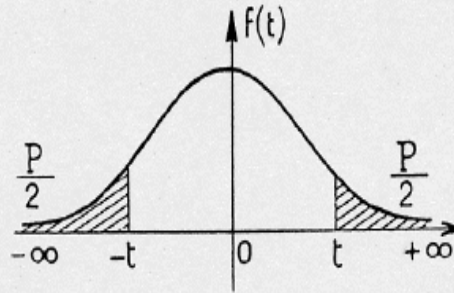


ν	aire relative										
	0,995	0,990	0,975	0,950	0,900	0,500	0,100	0,050	0,025	0,010	0,005
1	0.00004	0.00016	0.00098	0.00393	0.0158	0.455	2.71	3.84	5.02	6.63	7.88
2	0.0100	0.0201	0.0506	0.103	0.211	1.386	4.61	5.99	7.38	9.21	10.60
3	0.072	0.115	0.216	0.352	0.584	2.366	6.25	7.81	9.35	11.34	12.84
4	0.207	0.297	0.484	0.711	1.064	3.357	7.78	9.49	11.14	13.28	14.86
5	0.412	0.554	0.831	1.145	1.61	4.251	9.24	11.07	12.83	15.09	16.75
6	0.676	0.872	1.24	1.64	2.20	5.35	10.64	12.59	14.45	16.81	18.55
7	0.989	1.24	1.69	2.17	2.83	6.35	12.02	14.07	16.01	18.48	20.28
8	1.34	1.65	2.18	2.73	3.49	7.34	13.36	15.51	17.53	20.09	21.96
9	1.73	2.09	2.70	3.33	4.17	8.34	14.68	16.92	19.02	21.67	23.59
10	2.16	2.56	3.25	3.94	4.87	9.34	15.99	18.31	20.48	23.21	25.19
11	2.60	3.05	3.82	4.57	5.58	10.34	17.28	19.68	21.92	24.73	26.76
12	3.07	3.57	4.40	5.23	6.30	11.34	18.55	21.03	23.34	26.22	28.30
13	3.57	4.11	5.01	5.89	7.04	12.34	19.81	22.36	24.74	27.69	29.82
14	4.07	4.66	5.63	6.57	7.79	13.34	21.06	23.68	26.12	29.14	31.32
15	4.60	5.23	6.26	7.26	8.55	14.34	22.31	25.00	27.49	30.58	32.80
16	5.14	5.81	6.91	7.96	9.31	15.34	23.54	26.30	28.85	32.00	34.27
17	5.70	6.41	7.56	8.67	10.09	16.34	24.77	27.59	30.19	33.41	35.72
18	6.26	7.01	8.23	9.39	10.86	17.34	25.99	28.87	31.53	34.81	37.16
19	6.84	7.63	8.91	10.12	11.65	18.34	27.20	30.14	32.85	36.19	38.58
20	7.43	8.26	9.59	10.85	12.44	19.34	28.41	31.41	34.17	37.57	40.00
21	8.03	8.90	10.28	11.59	13.24	20.34	29.62	32.67	35.48	38.93	41.40
22	8.64	9.54	10.98	12.34	14.04	21.34	30.81	33.92	36.78	40.29	42.80
23	9.26	10.20	11.69	13.09	14.85	22.34	32.01	35.17	38.08	41.64	44.18
24	9.89	10.86	12.40	13.85	15.66	23.34	33.20	36.42	39.36	42.98	45.56
25	10.52	11.52	13.12	14.61	16.47	24.34	34.38	37.65	40.65	44.31	46.93
26	11.16	12.20	13.84	15.38	17.29	25.34	35.56	38.89	41.92	45.64	48.29
27	11.81	12.83	14.57	16.15	18.11	26.34	36.74	40.11	43.19	46.96	49.64
28	12.46	13.56	15.31	16.93	18.94	27.34	37.92	41.34	44.46	48.28	50.99
29	13.12	14.26	16.05	17.71	19.77	28.34	39.09	42.56	45.72	49.59	52.34
30	13.79	14.95	16.79	18.49	20.60	29.34	40.26	43.77	46.98	50.89	53.67
40	20.71	22.16	24.43	26.51	29.05	39.34	51.81	55.76	59.34	63.69	66.77
50	27.99	29.71	32.36	34.76	37.69	49.33	63.17	67.50	71.42	76.15	79.49
60	35.53	37.43	40.48	43.19	46.46	59.33	74.40	79.08	83.30	88.38	91.95
70	43.28	45.44	48.76	51.74	55.33	69.33	85.53	90.53	95.02	100.4	104.2
80	51.17	53.54	51.17	60.39	64.28	79.33	98.58	101.9	106.6	112.3	116.3
90	59.20	61.75	65.65	69.13	73.29	89.33	107.6	113.1	118.1	124.1	128.3
100	67.33	70.06	74.22	77.93	82.36	99.33	118.5	124.3	129.6	135.8	140.2

Annexe IV. Table de distribution Kolmogorov smirnov

n	P = .80	P = .90	P = .95	P = .98	P = .99
1	.90000	.95000	.97500	.99000	.99500
2	.68377	.77639	.84189	.90000	.92929
3	.56481	.63004	.70760	.78456	.82900
4	.49265	.56522	.62394	.68887	.73424
5	.44698	.50945	.56328	.62718	.66853
6	.41037	.46799	.51926	.57741	.61661
7	.38148	.43807	.48342	.53844	.57581
8	.35831	.40962	.45427	.50654	.54179
9	.33910	.38746	.43001	.47960	.51332
10	.32260	.36866	.40925	.45602	.48893
11	.30829	.35242	.39122	.43670	.46770
12	.29577	.33815	.37543	.41918	.44905
13	.28470	.32549	.36143	.40362	.43247
14	.27481	.31417	.34890	.38970	.41762
15	.26588	.30397	.33760	.37713	.40420
16	.25778	.29472	.32733	.36571	.39201
17	.25030	.28627	.31796	.35528	.38086
18	.24360	.27851	.30936	.34569	.37062
19	.23735	.27136	.30143	.33685	.36117
20	.23156	.26473	.29408	.32866	.35241
21	.22617	.25858	.28724	.32104	.34427
22	.22115	.25283	.28087	.31394	.33666
23	.21645	.24746	.27490	.30728	.32954
24	.21205	.24242	.26931	.30104	.32286
25	.20790	.23768	.26404	.29516	.31657
26	.20399	.23320	.25907	.28962	.31064
27	.20030	.22898	.25438	.28438	.30502
28	.19680	.22497	.24993	.27942	.29971
29	.19348	.22117	.24571	.27471	.29466
30	.19032	.21756	.24170	.27023	.28987
31	.18732	.21412	.23788	.26596	.28530
32	.18445	.21085	.23424	.26189	.28094
33	.18171	.20771	.23076	.25801	.27677
34	.17909	.20472	.22743	.25429	.27279
35	.17659	.20185	.22425	.25073	.26897
36	.17418	.19910	.22119	.24732	.26532
37	.17188	.19646	.21826	.24404	.26180
38	.16966	.19392	.21544	.24089	.25843
39	.16753	.19148	.21273	.23786	.25518
40	.16547	.18913	.21012	.23494	.25205
41	.16349	.18687	.20760	.23213	.24904
42	.16158	.18468	.20517	.22941	.24613
43	.15974	.18257	.20283	.22679	.24332
44	.15796	.18053	.20056	.22426	.24060
45	.15623	.17856	.19837	.22181	.23798
46	.15457	.17665	.19625	.21944	.23544
47	.15295	.17481	.19420	.21715	.23298
48	.15139	.17302	.19221	.21493	.23059
49	.14987	.17128	.19028	.21277	.22828
50	.14840	.16959	.18841	.21068	.22604
51	.14697	.16796	.18659	.20864	.22386
52	.14558	.16637	.18482	.20667	.22174
53	.14423	.16483	.18311	.20475	.21968
54	.14292	.16332	.18144	.20289	.21768
55	.14164	.16186	.17981	.20107	.21574
56	.14040	.16044	.17823	.19930	.21384
57	.13919	.15906	.17669	.19758	.21199
58	.13801	.15771	.17519	.19590	.21019
59	.13686	.15639	.17373	.19427	.20844
60	.13573	.15511	.17231	.19267	.20673
61	.13464	.15385	.17091	.19112	.20506
62	.13357	.15263	.16956	.18960	.20343
63	.13253	.15144	.16823	.18812	.20184
64	.13151	.15027	.16693	.18667	.20029
65	.13052	.14913	.16567	.18525	.19877
66	.12954	.14802	.16443	.18387	.19729
67	.12859	.14693	.16322	.18252	.19584
68	.12766	.14587	.16204	.18119	.19442
69	.12675	.14483	.16088	.17990	.19303
70	.12586	.14381	.15975	.17863	.19167
71	.12499	.14281	.15864	.17739	.19034
72	.12413	.14183	.15755	.17618	.18903
73	.12329	.14087	.15649	.17498	.18776
74	.12247	.13993	.15544	.17382	.18650
75	.12167	.13901	.15442	.17268	.18528
76	.12088	.13811	.15342	.17155	.18408
77	.12011	.13723	.15244	.17045	.18290
78	.11935	.13638	.15147	.16938	.18174
79	.11860	.13551	.15052	.16832	.18060
80	.11787	.13467	.14960	.16728	.17949
81	.11716	.13385	.14868	.16626	.17840
82	.11645	.13305	.14779	.16526	.17732
83	.11576	.13226	.14691	.16428	.17627
84	.11508	.13148	.14605	.16331	.17523
85	.11442	.13072	.14520	.16236	.17421
86	.11376	.12997	.14437	.16143	.17321
87	.11311	.12923	.14355	.16051	.17223
88	.11248	.12850	.14274	.15961	.17126
89	.11186	.12770	.14195	.15873	.17031
90	.11125	.12709	.14117	.15786	.16938
91	.11064	.12640	.14040	.15700	.16846
92	.11005	.12572	.13965	.15618	.16755
93	.10947	.12506	.13891	.15533	.16666
94	.10889	.12440	.13818	.15451	.16579
95	.10833	.12375	.13746	.15371	.16493
96	.10777	.12312	.13675	.15291	.16408
97	.10722	.12249	.13606	.15214	.16324
98	.10668	.12187	.13537	.15137	.16242
99	.10615	.12126	.13469	.15061	.16161
100	.10563	.12067	.13403	.14987	.16081
n > 100	$1.073/\sqrt{n}$	$1.223/\sqrt{n}$	$1.358/\sqrt{n}$	$1.518/\sqrt{n}$	$1.629/\sqrt{n}$

Annexe V. Table de Student



$\frac{P}{v}$	0,90	0,80	0,70	0,60	0,50	0,40	0,30	0,20	0,10	0,05	0,02	0,01	0,001
1	0,158	0,325	0,510	0,727	1,000	1,376	1,963	3,078	6,314	12,706	31,821	63,657	636,619
2	0,142	0,289	0,445	0,617	0,816	1,061	1,386	1,886	2,920	4,303	6,965	9,925	31,598
3	0,137	0,277	0,424	0,584	0,765	0,978	1,250	1,638	2,353	3,182	4,541	5,841	12,929
4	0,134	0,271	0,414	0,569	0,741	0,941	1,190	1,533	2,132	2,776	3,747	4,604	8,611
5	0,132	0,267	0,408	0,559	0,727	0,920	1,156	1,476	2,015	2,571	3,365	4,032	6,869
6	0,131	0,265	0,404	0,553	0,718	0,906	1,134	1,440	1,943	2,447	3,143	3,707	5,959
7	0,130	0,263	0,402	0,549	0,711	0,896	1,119	1,415	1,895	2,365	2,998	3,499	5,408
8	0,130	0,262	0,399	0,546	0,706	0,889	1,108	1,397	1,860	2,306	2,896	3,355	5,041
9	0,129	0,261	0,398	0,543	0,703	0,883	1,100	1,383	1,833	2,262	2,821	3,250	4,781
10	0,129	0,260	0,397	0,542	0,700	0,879	1,093	1,372	1,812	2,228	2,764	3,169	4,587
11	0,129	0,260	0,396	0,540	0,697	0,876	1,088	1,363	1,796	2,201	2,718	3,106	4,437
12	0,128	0,259	0,395	0,539	0,695	0,873	1,083	1,356	1,782	2,179	2,681	3,055	4,318
13	0,128	0,259	0,394	0,538	0,694	0,870	1,079	1,350	1,771	2,160	2,650	3,012	4,221
14	0,128	0,258	0,393	0,537	0,692	0,868	1,076	1,345	1,761	2,145	2,624	2,977	4,140
15	0,128	0,258	0,393	0,536	0,691	0,866	1,074	1,341	1,753	2,131	2,602	2,947	4,073
16	0,128	0,258	0,392	0,535	0,690	0,865	1,071	1,337	1,746	2,120	2,583	2,921	4,015
17	0,128	0,257	0,392	0,534	0,689	0,863	1,069	1,333	1,740	2,110	2,567	2,898	3,965
18	0,127	0,257	0,392	0,534	0,688	0,862	1,067	1,330	1,734	2,101	2,552	2,878	3,922
19	0,127	0,257	0,391	0,533	0,688	0,861	1,066	1,328	1,729	2,093	2,539	2,861	3,883
20	0,127	0,257	0,391	0,533	0,687	0,860	1,064	1,325	1,725	2,086	2,528	2,845	3,850
21	0,127	0,257	0,391	0,532	0,686	0,859	1,063	1,323	1,721	2,080	2,518	2,831	3,819
22	0,127	0,256	0,390	0,532	0,686	0,858	1,061	1,321	1,717	2,074	2,508	2,819	3,792
23	0,127	0,256	0,390	0,532	0,685	0,858	1,060	1,319	1,714	2,069	2,500	2,807	3,767
24	0,127	0,256	0,390	0,531	0,685	0,857	1,059	1,318	1,711	2,064	2,492	2,797	3,745
25	0,127	0,256	0,390	0,531	0,684	0,856	1,058	1,316	1,708	2,060	2,485	2,787	3,725
26	0,127	0,256	0,390	0,531	0,684	0,856	1,058	1,315	1,706	2,056	2,479	2,779	3,707
27	0,127	0,256	0,389	0,531	0,684	0,855	1,057	1,314	1,703	2,052	2,473	2,771	3,690
28	0,127	0,256	0,389	0,530	0,683	0,855	1,056	1,313	1,701	2,048	2,467	2,763	3,674
29	0,127	0,256	0,389	0,530	0,683	0,854	1,055	1,311	1,699	2,045	2,462	2,756	3,659
30	0,127	0,256	0,389	0,530	0,683	0,854	1,055	1,310	1,697	2,042	2,457	2,750	3,646
40	0,126	0,255	0,388	0,529	0,681	0,851	1,050	1,303	1,684	2,021	2,423	2,704	3,551
80	0,126	0,254	0,387	0,527	0,679	0,848	1,046	1,296	1,671	2,000	2,390	2,660	3,460
120	0,126	0,254	0,386	0,526	0,677	0,845	1,041	1,289	1,658	1,980	2,358	2,617	3,373
∞	0,126	0,253	0,385	0,524	0,674	0,842	1,036	1,282	1,645	1,960	2,326	2,576	3,291