

MÉMOIRE DE MAÎTRISE
INRS-EAU

MODÉLISATION DE VARIABLES DE PRÉCIPITATION
PAR DES DISTRIBUTIONS STATISTIQUES

APPLICATION À LA RÉGION DU QUÉBEC ET DU NOUVEAU-BRUNSWICK

PAR
MARIO HACHÉ

UNIVERSITÉ DU QUÉBEC
2800, rue Einstein, Suite 205
Ste-Foy, QC
G1V 4C7

JUILLET 1993

REMERCIEMENTS

Je souhaite exprimer toute ma reconnaissance à monsieur Bernard Bobée, mon directeur, ainsi qu'à monsieur Fahim Ashkar, mon co-directeur. Leurs commentaires et leurs conseils ont guidé mon travail et ma démarche scientifique tout au long de la maîtrise.

Je tiens aussi à remercier monsieur Jacques Bernier pour ses appréciations et commentaires qui ont été très utiles lors de la rédaction du présent mémoire. Également, la collaboration de l'équipe de recherche de monsieur Bobée, en particulier celle de monsieur Luc Perreault, a été grandement appréciée.

Finalement, je voudrais faire une mention spéciale à ma compagne, Christine, qui par ses nombreux encouragements et sa compréhension a contribué à la réussite de mon projet, de même qu'à mes deux enfants, Annie et Pascal, pour l'émerveillement constant qu'ils m'ont apporté.

RÉSUMÉ

L'étude statistique des phénomènes hydrométéorologiques occupe une place importante dans plusieurs domaines théoriques et pratiques. Par exemple, la modélisation statistique des variables de précipitation est utile dans la planification de la ressource en eau, le dimensionnement des ouvrages hydrauliques et la lutte contre les inondations.

Cette étude a pour but d'identifier la distribution statistique la plus adéquate pour représenter les variables de précipitation suivantes:

- durées des épisodes secs et pluvieux (et les cycles);
- hauteur de précipitation journalière (et mensuelle);
- maximum annuel de précipitation.

Différents tests et critères statistiques tels le test du chi-2, le test du rapport des vraisemblances maximales et le critère d'information d'Akaike ont permis d'identifier la distribution la plus appropriée pour chaque variable.

Les principales conclusions de cette étude sont les suivantes:

- La distribution géométrique représente le mieux la durée des épisodes secs alors que la distribution binomiale négative tronquée représente la mieux la durée des épisodes pluvieux. La combinaison des lois binomiale négative tronquée et géométrique a conduit aux meilleurs résultats pour les cycles.
- La distribution exponentielle mixte a conduit aux meilleurs résultats en ce qui concerne les hauteurs de pluie journalière. Les distributions gamma et Weibull représentent le mieux les hauteurs de pluie mensuelle.
- Les valeurs maximums de précipitation sont représentées adéquatement par les distributions log-normale, log-gamma et Gumbel.

TABLE DES MATIÈRES

REMERCIEMENTS	i
RÉSUMÉ	ii
TABLE DES MATIÈRES	iii
LISTE DES TABLEAUX	vii
LISTE DES FIGURES	x
PRINCIPALES NOTATIONS	xii
1. INTRODUCTION	1
1.1 Problématique	1
1.2 Objectif et méthodologie	2
2. DISTRIBUTIONS STATISTIQUES	4
2.1 Considérations générales	4
2.1.1 Modélisation d'un échantillon par une distribution statistique	4
2.1.2 Définition d'une distribution statistique	4
2.1.3 Moments et coefficients	5
a) Moments d'une distribution	5
b) Moments d'un échantillon	6
2.2 Estimation des paramètres	7
2.2.1 Propriétés des estimateurs	7
a) Estimateurs convergents	8
b) Estimateurs non-biaisés	8
c) Estimateurs efficaces	8
d) Statistique exhaustive	9
2.2.2 Méthodes d'estimation des paramètres	9
a) Maximum de vraisemblance (MV)	9
b) Méthode des moments (MM)	10
c) Méthode des moments pondérés (MMP)	11
d) Méthode généralisée des moments	11
2.3 Distributions statistiques	12
2.3.1 Loi géométrique	12
a) Estimation du paramètre q	12
2.3.2 Loi "série logarithmique"	12
a) Estimation du paramètre θ	13
2.3.3 Loi binomiale négative tronquée	13
a) Estimation des paramètres par la méthode des moments	14
b) Estimation des paramètres par la méthode du maximum de vraisemblance	15
2.3.4 Loi normale	16
2.3.5 Loi log-normale (2 paramètres)	16

	a)	Estimation des paramètres par la méthode des moments	17
	b)	Estimation des paramètres par la méthode du maximum de vraisemblance	17
2.3.6		Loi log-normale à 3 paramètres	18
	a)	Estimation des paramètres par la méthode des moments	18
	b)	Estimation des paramètres par la méthode du maximum de vraisemblance	19
2.3.7		Loi Weibull	20
2.3.8		Loi Gumbel	21
2.3.9		Loi généralisée des valeurs extrêmes (GEV)	22
2.3.10		Loi gamma	23
2.3.11		Loi gamma généralisée	24
2.3.12		Loi Pearson type 3	25
2.3.13		Loi log-Pearson type 3	25
2.3.14		Loi log-gamma	26
2.3.15		Loi exponentielle	27
	a)	Estimation du paramètre α	27
2.3.16		Loi exponentielle mixte	28
	a)	Estimation des paramètres par la méthode des moments	28
	b)	Estimation des paramètres par la méthode du maximum de vraisemblance	29
2.3.17		Loi Kappa (2 paramètres)	30
	a)	Estimation des paramètres par la méthode des moments	30
	b)	Estimation des paramètres par la méthode du maximum de vraisemblance	31
3.		THÉORIE DE L'AJUSTEMENT	32
3.1		Critères statistiques	32
	3.1.1	Indépendance	32
	3.1.2	Stationnarité	33
	3.1.3	Homogénéité	33
	3.1.4	Valeurs singulières	34
3.2		Tests statistiques	34
	3.2.1	Test de Wald-Wolfowitz	35
	3.2.2	Test de Mann-Withney	36
	3.2.3	Test de Kruskal-Wallis	38
	3.2.4	Test de Grubbs-Beck	39
3.3		Critères d'adéquation	39
	3.3.1	Test du chi-2	40
	3.3.2	Test du rapport des vraisemblances maximales	41
	3.3.3	Critère d'information d'Akaike	42

3.3.4	Représentation graphique	43
4.	REVUE DE LITTÉRATURE	45
4.1	Variables de précipitation	45
4.2	Applications pratiques	45
4.2.1	Variables de base	45
4.2.2	Valeurs extrêmes de précipitation	48
4.3	Occurrence des précipitations	50
4.4	Durée des épisodes pluvieux et secs	51
4.5	Hauteur de précipitation	53
4.6	Valeurs extrêmes de précipitation	56
4.6.1	Séries annuelles de maxima	57
4.6.2	Séries de durée partielle	59
4.7	Autres phénomènes météorologiques	60
5.	ÉPISODES PLUVIEUX ET SECS	62
5.1	Choix des stations	62
5.2	Définition des épisodes	63
5.3	Critères et tests statistiques	64
5.3.1	Indépendance	64
5.3.2	Homogénéité	67
5.3.3	Valeurs singulières	68
5.4	Caractéristiques statistiques des épisodes pluvieux et secs	69
5.5	Ajustement des épisodes pluvieux et secs	70
5.5.1	Distributions et méthodes d'estimation considérées	70
5.5.2	Calculs effectués	70
5.6	Lien entre la chaîne de Markov et la loi géométrique	72
5.7	Test du chi-2	75
5.8	Critère d'information d'Akaike	77
5.9	Les cycles	79
5.9.1	Lois de probabilité des cycles	79
5.9.2	Test du chi-2	81
5.9.3	Critère d'information d'Akaike	82
5.10	Conclusion	83
6.	HAUTEURS DE PRÉCIPITATION JOURNALIÈRE	85
6.1	Choix des stations	85
6.2	Critères et tests statistiques	85
6.2.1	Indépendance	85
6.2.2	Homogénéité	87
6.2.3	Valeurs singulières	87
6.3	Caractéristiques statistiques des hauteurs de pluie	88
6.4	Ajustement des hauteurs de pluie journalière	89
6.4.1	Distributions et méthodes d'estimation considérées	89
6.4.2	Calculs effectués	90

6.5	Test du chi-2	91
6.6	Critère d'information d'Akaïke	93
6.7	Hauteurs de précipitation annuelle et mensuelle	95
6.8	Conclusion	99
7.	VALEURS EXTRÊMES DE PRÉCIPITATION	101
7.1	Choix des stations	101
7.2	Critères et tests statistiques	103
	7.2.1 Indépendance	103
	7.2.2 Homogénéité	104
	7.2.3 Valeurs singulières	106
7.3	Caractéristiques statistiques des extrêmes de précipitation	107
7.4	Ajustement des séries annuelles de maxima	109
	7.4.1 Distributions et méthodes d'estimation considérées	109
	7.4.2 Calculs effectués	110
7.5	Comparaison entre les valeurs observées et théoriques	111
	7.5.1 Comparaison des méthodes d'ajustement	114
	7.5.2 Comparaison des distributions	123
7.6	Critère d'information d'Akaïke	126
7.7	Test du rapport des vraisemblances maximales	130
7.8	Conclusion	131
8.	CONCLUSION GÉNÉRALE ET DISCUSSION	134
	RÉFÉRENCES	137
	ANNEXES	144
A.	Origine du critère d'information d'Akaïke	144
B.	Distribution série logarithmique: cas limite de la loi binomiale négative tronquée	147
C.	Changement d'unité de mesure avec la loi log-gamma	149
D.	Exemples d'ajustement sur papier de probabilité	151

LISTE DES TABLEAUX

Tableau 4.1	Période de retour recommandée au Québec pour les bassins de 20 km ² et moins	50
Tableau 4.2	Distributions utilisées pour modéliser les durées des épisodes pluvieux et secs	52
Tableau 4.3	Distributions utilisées pour modéliser les hauteurs de précipitation	55
Tableau 4.4	Distributions utilisées pour modéliser les séries annuelles de maxima	58
Tableau 4.5	Distributions utilisées pour modéliser les séries de durée partielle	59
Tableau 4.6	Facteurs de conversion entre les résultats basés sur les séries annuelles de maxima et ceux basés sur les séries de durée partielle	60
Tableau 5.1	Station sélectionnées pour l'étude des durées des épisodes pluvieux et secs	63
Tableau 5.2ab	La statistique $ u $ du test d'indépendance de Wald-Wolfowitz pour les épisodes pluvieux (a) et secs (b) (rejet à 5%)	65
Tableau 5.3	La statistique τ du test d'homogénéité mensuelle de Kruskal-Wallis pour les épisodes pluvieux et secs	68
Tableau 5.4ab	Caractéristiques statistiques des épisodes pluvieux (a) et secs (b)	69
Tableau 5.5ab	Paramètres estimés pour les épisodes pluvieux (a) et secs (b)	72
Tableau 5.6ab	Valeurs du chi-2 pour les épisodes pluvieux (a) et secs (b)	76
Tableau 5.7	P-value du chi-2 pour les épisodes pluvieux et secs	77
Tableau 5.8	Valeurs du CIA pour les épisodes pluvieux et secs	78
Tableau 5.9	Valeurs du chi-2 pour les cycles	82

Tableau 5.10	Valeurs du CIA pour les cycles	82
Tableau 6.1	La statistique $ u $ du test d'indépendance de Wald-Wolfowitz pour les hauteurs de pluie journalière	85
Tableau 6.2	La statistique τ du test d'homogénéité mensuelle de Kruskal-Wallis pour les hauteurs de pluie journalière	87
Tableau 6.3	Caractéristiques statistiques des hauteurs de pluie journalière	88
Tableau 6.4	Paramètres estimés pour les hauteurs de pluie journalière . . .	92
Tableau 6.5	Valeurs du chi-2 pour les hauteurs de pluie journalière	93
Tableau 6.6	Valeurs du CIA pour les hauteurs de pluie journalière	94
Tableau 6.7	Stations choisies pour l'étude des hauteurs de pluie annuelle et mensuelle	95
Tableau 6.8	Valeurs du CIA pour les hauteurs de pluie annuelle et mensuelle	97
Tableau 7.1	Stations sélectionnées pour l'étude des séries annuelles de maxima	102
Tableau 7.2	La statistique $ u $ du test d'indépendance de Wald-Wolfowitz pour les séries annuelles de maxima	104
Tableau 7.3	La statistique $ u $ du test d'homogénéité de Mann-Whitney pour les séries annuelles de maxima	106
Tableau 7.4	Résultats du test de détection des valeurs singulières de Grubbs-Beck pour les séries annuelles de maxima	106
Tableau 7.5	Caractéristiques statistiques des séries annuelles de maxima	108
Tableau 7.6	Hauteurs de pluie interpolées $O(T)$ entre les observations adjacentes	113
Tableau 7.7abc	Moyennes des écarts absolus $A_{D,M}$	115-117
Tableau 7.8abc	Rangs des la moyenne des écarts absolue $A_{D,M}$	119-121

Tableau 7.9	Moyennes des écarts absolus pour les méthodes sélectionnées	124
Tableau 7.10	Rangs de la moyenne des écarts absolus pour les méthodes sélectionnées	125
Tableau 7.11	Valeurs du CIA pour les séries annuelles de maxima	128
Tableau 7.12	Rangs de la valeur du CIA	129
Tableau 8.11	Synthèse de l'étude de représentation des variables de précipitation par des lois statistiques	135

LISTE DES FIGURES

Figure 4.1	Hauteurs hypothétiques de pluie journalière au cours d'un mois	46
Figure 5.1	Boîte à moustaches de l'autocorrélation d'ordre un des durées des épisodes pluvieux et secs	66
Figure 5.2	Histogramme typique de la durée des épisodes secs	71
Figure 6.1	Boîtes à moustaches de l'autocorrélation d'ordre un des hauteurs de pluie journalière	86
Figure 6.2	Histogramme typique de la hauteur de pluie journalière	89
Figure D.1	Ajustement sur papier de probabilité des hauteurs de pluie journalière - station QU1 (EX,WE,EXM)	152
Figure D.2	Ajustement sur papier de probabilité des hauteurs de pluie journalière - station QU3 (EX,WE,EXM)	153
Figure D.3	Ajustement sur papier de probabilité des hauteurs de pluie journalière - station QU4 (EX,WE,EXM)	154
Figure D.4	Ajustement sur papier de probabilité des hauteurs de pluie journalière - station QU5 (EX,WE,EXM)	155
Figure D.5	Ajustement sur papier de probabilité des hauteurs de pluie journalière - station NB2 (EX,WE,EXM)	156
Figure D.6	Ajustement sur papier de probabilité des hauteurs de pluie journalière - station NB3 (EX,WE,EXM)	157
Figure D.7	Ajustement sur papier de probabilité des séries annuelles de maxima - station QU5-1H (GU)	158
Figure D.8	Ajustement sur papier de probabilité des séries annuelles de maxima - station NE4-24H (LG)	159
Figure D.9	Ajustement sur papier de probabilité des séries annuelles de maxima - station NB3-5M (GG)	160
Figure D.10	Ajustement sur papier de probabilité des séries annuelles de maxima - station TN1-24H (GEV)	161

Figure D.11	Ajustement sur papier de probabilité des séries annuelles de maxima - station QU3-24H (LP3)	162
Figure D.12	Ajustement sur papier de probabilité des séries annuelles de maxima - station QU3-24H (GU,GA,LG)	163
Figure D.13	Ajustement sur papier de probabilité des séries annuelles de maxima - station QU3-24H (GG,P3,GEV,LP3)	164

PRINCIPALES NOTATIONS

$A_{D,M}$	=	moyenne des écarts absolus
CIA	=	critère d'information d'Akaïke
$B(.,.)$	=	fonction bêta
BN	=	distribution binomiale négative
BNT	=	distribution binomiale négative tronquée
C_k	=	coefficient d'aplatissement de la population
C_s	=	coefficient d'asymétrie de la population
C_v	=	coefficient de variation de la population
\hat{C}_k	=	coefficient d'aplatissement de l'échantillon
\hat{C}_s	=	coefficient d'asymétrie de l'échantillon
\hat{C}_v	=	coefficient de variation de l'échantillon
$E[.]$	=	espérance mathématique
$E_{D,M}$	=	écart relatif en pourcentage
EX	=	distribution exponentielle
EXM	=	distribution exponentielle mixte
$f(.)$	=	fonction de densité
$F(.)$	=	fonction de distribution
f.d.p	=	fonction densité de probabilité
GA	=	distribution gamma
GEO	=	distribution géométrique
GEV	=	loi généralisée des valeurs extrêmes
GG	=	distribution gamma généralisée
GT	=	distribution Grassia-Tadikamalla
GU	=	distribution Gumbel
H	=	hauteur de pluie
H_0	=	hypothèse nulle
H_1	=	hypothèse alternative
i.i.d.	=	indépendant et identiquement distribué

KA	=	distribution Kappa à 2 paramètres
KA3	=	distribution Kappa à 3 paramètres
L	=	fonction de vraisemblance
LF	=	loi de fuite
LGU	=	distribution log-Gumbel
LN	=	distribution log-normale à 2 paramètres
LN3	=	distribution log-normale à 3 paramètres
lnL	=	logarithme de la fonction de vraisemblance
LP3	=	distribution log-Pearson type 3
m'_r	=	moment non-centré d'ordre r de l'échantillon
m_r	=	moment centré par rapport à la moyenne d'ordre r de l'échantillon
MM	=	méthode des moments
MM1	=	méthode généralisée des moments (0,1,2)
MMP	=	méthode des moments pondérés
MOY	=	moyenne
MV	=	méthode du maximum de vraisemblance
n	=	taille de l'échantillon
N	=	distribution normale
O(T)	=	hauteur de pluie interpolée de période de retour T
P[A]	=	probabilité que l'événement A arrive
P3	=	distribution Pearson type 3
$R_{D,M}$	=	rang de $S_{D,M}$
s	=	écart-type de l'échantillon
s^2	=	variance non-biaisée de l'échantillon
$S_{D,M}$	=	somme des rangs de $A_{D,M}$
SAM	=	méthode généralisée des moments (-1,0,1)
SEA	=	Service de l'Environnement Atmosphérique
SL	=	distribution série logarithmique
T	=	période de retour
T_p	=	épisode pluvieux

T_s	=	épisode sec
WE	=	distribution Weibull à 2 paramètres
x_i	=	élément i de l'échantillon
x_T	=	événement de période de retour T
\bar{x}	=	moyenne de l'échantillon
X	=	variable aléatoire
$\Gamma(.)$	=	fonction gamma
ρ	=	coefficient de corrélation
ρ_1	=	autocorrélation d'ordre un
μ_r'	=	moment non-centré d'ordre r de la distribution
μ_r	=	moment centré par rapport à la moyenne d'ordre r de la distribution
μ_1'	=	moyenne de la distribution
μ_2	=	variance de la distribution
$\chi^2_{\nu, \alpha}$	=	variable chi-2 à ν degrés de liberté dont la probabilité au dépassement est α

CHAPITRE 1 INTRODUCTION

1.1 PROBLÉMATIQUE

Les précipitations (neige et pluie) sont abondantes au Québec et dans les Provinces Atlantiques. Au Québec, elles varient entre 900 mm par année dans la région de Montréal à plus de 1400 mm sur les sommets des Laurentides pour décroître à nouveau vers le nord où il tombe moins de 400 mm. Dans les Provinces Atlantiques, le relief et l'orientation des terres par rapport à la mer influencent la fréquence et la quantité des précipitations. Les hautes-terres reçoivent plus de 1400 mm par an alors qu'il en tombe environ 1000 mm dans les régions côtières.

Il est évident que d'aussi grandes quantités d'eau influencent plusieurs aspects significatifs de notre vie et de notre environnement. Par exemple, le dimensionnement des ouvrages hydrauliques dépend souvent de critères reliés aux précipitations. La gestion des réservoirs ainsi que l'opération des systèmes de drainage urbain exigent également une bonne connaissance du régime des précipitations. De plus, les caractéristiques fondamentales des séries pluviométriques sont utiles dans le domaine agricole et la prévention des feux de forêts. Il est également important de connaître les précipitations afin de protéger les basses-terres contre les inondations.

La condensation massive de vapeur d'eau provenant de l'évaporation au sol et des nappes d'eau et de l'évapotranspiration des plantes provoque les précipitations. En général, ce processus peut se produire de 4 manières:

- Un écoulement d'air convergent dans une région restreinte;
- Une masse d'air forcé à remonter une barrière topographique;
- Une convection naturelle de l'air;
- Une tempête cyclonique (convergence à grande échelle de masses d'air).

Les nombreux mécanismes entraînant les précipitations sont gouvernés par

des lois physiques et des phénomènes atmosphériques très complexes. Bien qu'il soit possible de les exprimer sous forme d'équations, les relations internes entre ces divers mécanismes rendent la modélisation des précipitations en termes mathématiques extrêmement ardue. Il existe cependant un moyen plus simple d'extraire l'information contenue dans les séries pluviométriques. Il s'agit de modéliser les variables de précipitation par des distributions statistiques. Il est alors possible d'estimer la probabilité au dépassement des niveaux de précipitation, ce qui a une grande importance en pratique.

Un grand nombre de distributions statistiques et de méthodes d'estimation des paramètres sont disponibles et ont été utilisées pour effectuer l'étude des précipitations. Il n'existe pas de consensus face au choix de la distribution à utiliser pour chacune des variables de précipitation. Les lois statistiques utilisées varient suivant les pays et d'un chercheur ou d'un utilisateur à l'autre. L'emploi d'une distribution plutôt qu'une autre peut cependant changer considérablement les conclusions et les conséquences d'une analyse. Par exemple, une sur-estimation de la hauteur d'un barrage basée sur une sur-estimation de la quantité de précipitations entraîne des coûts supplémentaires alors qu'une sous-estimation se traduit par des dégâts matériels et d'éventuelles pertes de vies. Il est donc essentiel de représenter les séries pluviométriques par la meilleure distribution possible.

1.2 OBJECTIF ET MÉTHODOLOGIE

Face à la multitude de lois statistiques qui s'offre à nous, une question nous vient à l'esprit: "Quelle est la distribution statistique la plus adéquate dans une région donnée pour une certaine variable?". L'objectif principal de notre étude est de répondre à cette question. C'est à dire de déterminer la distribution la plus adéquate pour différentes variables de précipitations. L'analyse porte plus particulièrement sur les variables suivantes:

- la hauteur de pluie journalière;
- la durée des épisodes pluvieux et secs;

- le maximum annuel de précipitation.

À cette liste s'ajoute la durée des cycles et la hauteur des précipitations mensuelles et annuelles qui sont traitées de manière moins détaillée.

L'atteinte de notre objectif se fait en 4 étapes qui sont habituellement suivies lors de l'ajustement des phénomènes physiques par des distributions statistiques:

- a) La sélection d'un échantillon pour lequel les données répondent à certains critères statistiques;
- b) La sélection des distributions statistiques pouvant potentiellement représenter l'échantillon;
- c) Le choix des méthodes d'estimation des paramètres pour chaque distribution;
- d) La comparaison des ajustements par différentes méthodes.

Une description des distributions statistiques et des méthodes d'estimation considérées tout au long de l'étude est présentée au chapitre 2. La théorie relative aux points a) et d) est traitée en détail au chapitre 3. Une revue des principales distributions employées pour ajuster les variables considérés se trouve au chapitre 4. En dernier lieu, les chapitres 5, 6 et 7 sont consacrés à l'ajustement des variables mentionnées précédemment.

La majorité des ajustements et des tests sont effectués à l'aide du logiciel AJUSTE 1.1 (INRS-Eau, 1992).

CHAPITRE 2 DISTRIBUTIONS STATISTIQUES

Ce chapitre a pour objectif de décrire les principales caractéristiques des distributions statistiques et des méthodes d'estimation des paramètres qui sont employées dans cette étude. Un grand nombre de ces lois sont incluses dans le logiciel d'ajustement automatique AJUSTE 1.1 (INRS-Eau, 1992). Il s'agit des lois Normale, Weibull, Gumbel, GEV, gamma, gamma généralisée, log-gamma, Pearson type 3 et log-Pearson type 3. Il n'y aura donc qu'une description succincte de ces distributions statistiques. Nous donnerons davantage de détails sur les distributions n'étant pas incluses dans le logiciel AJUSTE 1.1 et pour lesquelles il a été nécessaire de programmer les méthodes d'estimation des paramètres en langage GAUSS.

2.1 CONSIDÉRATIONS GÉNÉRALES

2.1.1 Modélisation d'un échantillon par une distribution statistique

L'estimation des paramètres d'une distribution nécessite un échantillon de taille suffisante, constitué de réalisations indépendantes et identiquement distribuées (i.i.d) de la variable aléatoire considérée et qui soient représentatives de la population étudiée. Cette hypothèse (i.i.d.) est utilisée afin de quantifier l'incertitude de la valeur des paramètres et la qualité de l'ajustement. Ce modèle d'observations i.i.d. rend également possible la comparaison des différentes méthodes d'estimation, par exemple, en termes de leurs écarts quadratiques moyens. Il est donc important de vérifier l'indépendance, l'homogénéité, la stationnarité et les valeurs singulières d'un échantillon pour voir s'il répond à l'hypothèse i.i.d. Cet aspect de l'ajustement est traité en détail au chapitre 3.

2.1.2 Définition d'une distribution statistique

Avant d'aller plus loin, il est important de définir en termes mathématiques ce qu'est une distribution statistique. Soit une variable aléatoire continue X . La fonction $f(x)$ est une fonction densité de probabilité (f.d.p.) si:

$$1) f(x) \geq 0 \quad \text{pour } x \text{ réel} \qquad (2.1a)$$

$$2) \int_{-\infty}^{\infty} f(x) dx = 1 \quad (2.1b)$$

$$3) P[a \leq x \leq b] = \int_a^b f(x) dx \quad \text{pour } a \text{ et } b \text{ réel} \quad (2.1c)$$

La fonction de distribution $F(x)$ est définie par:

$$F(x) = P[X \leq x] \quad \text{pour } x \text{ réel} \quad (2.2)$$

et elle est reliée à la fonction densité de probabilité par la relation:

$$F(x) = \int_{-\infty}^x f(t) dt \quad (2.3)$$

2.1.3 Moments et coefficients

La forme d'une distribution est déterminée par la valeur de ses paramètres. Elle est également caractérisée par ses moments. En calculant les moments empiriques d'un échantillon et en les comparant avec les moments théoriques de diverses distributions, il est possible de sélectionner certaines distributions pour représenter l'échantillon et d'en écarter d'autres. D'un autre côté, l'estimation des paramètres par la méthode des moments est basée sur l'égalité entre les moments théoriques et ceux de l'échantillon.

a) Moments d'une distribution

Le moment non-centré d'ordre r d'une distribution $F(x)$ est défini par:

$$\mu'_r = E[X^r] = \int_{-\infty}^{\infty} x^r f(x) dx \quad (2.4a)$$

et le moment centré par rapport à la moyenne s'écrit:

$$\mu_r = E[(X - \mu'_1)^r] = \int_{-\infty}^{\infty} (x - \mu'_1)^r f(x) dx \quad (2.4b)$$

Le moment non-centré d'ordre 1 (μ'_1) et le moment centré d'ordre 2 (μ_2) sont

respectivement la moyenne et la variance de la distribution.

Trois coefficients sans dimension sont définis à partir de ces moments:

$$\text{coefficient de variation} \quad C_v = \mu_2^{1/2} / \mu_1 \quad (2.5a)$$

$$\text{coefficient d'asymétrie} \quad C_s = \mu_3 / \mu_2^{3/2} \quad (2.5b)$$

$$\text{coefficient d'aplatissement} \quad C_k = \mu_4 / \mu_2^2 \quad (2.5c)$$

Les coefficients C_s et C_k peuvent être comparés à 0 et 3 qui sont respectivement les coefficients d'asymétrie et d'aplatissement de la loi normale. Le signe de C_s indique si la distribution a une asymétrie à droite (positive, $C_s > 0$) ou à gauche (négative, $C_s < 0$). Lorsque $C_k < 3$ la distribution a un pic plus accentué que la distribution normale tandis que lorsque $C_k > 3$ la distribution a un pic moins prédominant que la distribution normale.

b) Moments d'un échantillon

Soit un échantillon x_1, \dots, x_n . Le moment non-centré et le moment centré par rapport à la moyenne d'ordre r de l'échantillon sont définis respectivement par:

$$m_r' = \frac{1}{n} \sum_{i=1}^n x_i^r \quad (2.6a)$$

$$m_r = \frac{1}{n} \sum_{i=1}^n (x_i - m_1')^r \quad (2.6b)$$

En particulier, m_1' est la moyenne de l'échantillon (souvent notée \bar{x}) et m_2 est la variance biaisée de l'échantillon. La variance non-biaisée est obtenue par:

$$s^2 = \frac{n}{(n-1)} m_2 \quad (2.7)$$

Les coefficients de variation, d'asymétrie et d'aplatissement peuvent être définis à partir des moments de l'échantillon. On peut tenir compte, dans une certaine mesure, du biais introduit par les moments m_r , qui sont des estimateurs

biaisés des moments μ_r . Les différents coefficients avec la correction usuelle du biais sont donnés par:

$$\text{coefficient de variation} \quad \hat{C}_v = \left(\frac{n}{n-1} \right)^{1/2} \frac{m_2^{1/2}}{m_1'} \quad (2.8a)$$

$$\text{coefficient d'asymétrie} \quad \hat{C}_s = \frac{[n(n-1)]^{1/2} m_3}{n-2 m_2^{3/2}} \quad (2.8b)$$

$$\text{coefficient d'aplatissement} \quad \hat{C}_k = \frac{n(n-1) m_4}{(n-2)(n-3) m_2^2} \quad (2.8c)$$

Il est important de mentionner que ces coefficients sont quand même biaisés malgré la correction apportée.

2.2 ESTIMATION DES PARAMÈTRES

Les distributions statistiques dépendent d'un certain nombre de paramètres. À moins d'être connue d'avance, la valeur des paramètres doit être estimée à partir d'un échantillon. Les propriétés souhaitables des estimateurs et les méthodes d'estimation utilisées dans cette étude sont brièvement décrites dans les sections suivantes.

2.2.1 Propriétés des estimateurs

Un bon estimateur $\hat{\theta}$ du paramètre θ doit satisfaire 4 propriétés: convergent, non-biaisé, efficace et suffisant. Un estimateur non-biaisé et efficace est souvent appelé le "meilleur" estimateur.

a) Estimateurs convergents

L'estimateur $\hat{\theta}$ est convergent (ou correct) lorsque celui-ci tend en probabilité vers la vraie valeur θ . En terme mathématique, on a que pour tout $\epsilon > 0$:

$$P[|\hat{\theta} - \theta| > \epsilon] \rightarrow 0 \text{ lorsque } n \rightarrow \infty \quad (2.9)$$

b) Estimateurs non-biaisés

Lorsque l'espérance mathématique de l'estimateur est égale à la vraie valeur du paramètre on dit que l'estimateur est non-biaisé:

$$E(\hat{\theta}) = \theta \quad (2.10)$$

Un estimateur correct n'est pas nécessairement non-biaisé.

c) Estimateurs efficaces

Lorsqu'il existe plusieurs estimateurs de $\hat{\theta}$, celui ayant la variance minimum est préféré car sa distribution autour de θ est moins dispersée. Lorsqu'il existe une valeur limite minimum de la variance, l'estimateur conduisant à cette limite est un estimateur efficace.

Généralement, la variance minimum est obtenue lorsque la taille de l'échantillon tend vers l'infini et dans ce cas on parle d'un estimateur asymptotiquement efficace.

L'efficacité relative de l'estimateur $\hat{\theta}_1$ par rapport à l'estimateur $\hat{\theta}_2$ est définie par:

$$EFF(\hat{\theta}_1/\hat{\theta}_2) = E[(\hat{\theta}_2 - \theta)^2] / E[(\hat{\theta}_1 - \theta)^2] \quad (2.11)$$

Ce qui est équivalent au rapport des variances lorsque les 2 estimateurs sont non-biaisés.

d) Statistique exhaustive

La statistique T est exhaustive si elle contient toute l'information de l'échantillon relative à θ . On montre que la statistique T est exhaustive si et seulement si la fonction de densité jointe peut être mise sous la forme:

$$\prod_{i=1}^n f(x_i) = g[T(x_1, \dots, x_n), \theta] h(x_1, \dots, x_n) \quad (2.12)$$

où $g[T(x_1, \dots, x_n), \theta]$ dépend uniquement de $T(x_1, \dots, x_n)$ et θ ;
 $h(x_1, \dots, x_n)$ est une fonction des observations et ne dépend pas de θ .

2.2.2 Méthodes d'estimation des paramètres

Il existe plusieurs méthodes pour estimer les paramètres d'une distribution statistique. Nous verrons tout d'abord les méthodes bien connues du maximum de vraisemblance et des moments. Par la suite, nous présenterons brièvement la méthode des moments pondérés et la méthode généralisée des moments.

a) Maximum de vraisemblance (MV)

La méthode du maximum de vraisemblance consiste, pour un échantillon donnée, à maximiser la fonction de vraisemblance (fonction de densité jointe) par rapport aux paramètres. Cette méthode est habituellement complexe mais a généralement des propriétés asymptotiques intéressantes.

Soit un échantillon aléatoire x_1, x_2, \dots, x_n tiré d'une distribution $F(x; \theta_1, \theta_2, \dots, \theta_p)$. Lorsqu'ils existent, les estimateurs obtenus par la méthode du maximum de vraisemblance sont les solutions $\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_p$ du système de p équations:

$$\frac{\partial L(\theta_1, \theta_2, \dots, \theta_p)}{\partial \theta_r} = 0 \quad r = 1, 2, \dots, p \quad (2.13)$$

où la fonction de vraisemblance est définie par:

$$L(\theta_1, \theta_2, \dots, \theta_p) = \prod_{i=1}^n f(x_i, \theta_1, \theta_2, \dots, \theta_p) \quad (2.14)$$

Il est souvent plus simple de maximiser le logarithme naturel de la fonction de vraisemblance que la vraisemblance elle-même. Une ou l'autre méthode conduit au même maximum car la fonction logarithmique est une fonction monotone croissante.

Dans le cas où il existe des statistiques conjointement exhaustives pour les paramètres, les estimateurs du maximum de vraisemblance sont fonction de ces

statistiques exhaustives. Lorsqu'il n'existe pas de statistiques exhaustives pour les paramètres, la théorie des grands échantillons admet sous certaines conditions des propriétés intéressantes pour les estimateurs obtenus par la méthode du maximum de vraisemblance. Plus précisément, ces estimateurs sont corrects, asymptotiquement non-biaisés et asymptotiquement efficaces.

b) Méthode des moments (MM)

La méthode des moments, appliquée à une loi à p paramètres, consiste à évaluer p moments théoriques indépendants de la distribution à p moments correspondants de l'échantillon. Cette méthode est généralement simple et les estimateurs de la méthode des moments peuvent être utilisés comme valeurs initiales dans la méthode du maximum de vraisemblance lorsque celle-ci nécessite une résolution par processus itératif.

Soit un échantillon aléatoire x_1, x_2, \dots, x_n tiré d'une distribution $F(x; \theta_1, \theta_2, \dots, \theta_p)$. Les estimateurs obtenus par la méthode des moments sont les solutions $\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_p$ du système de p équations:

$$\mu'_r = m'_r \quad r = 1, 2, \dots, p \quad (2.15)$$

où μ'_r et m'_r sont les moments non-centrés d'ordre r de la distribution et de l'échantillon définis aux équations (2.4a) et (2.6a).

De la même manière que pour les moments non-centrés, les estimateurs de la méthode des moments peuvent être obtenus en utilisant les moments centrés ou des coefficients (variation, d'asymétrie).

Les estimations obtenues par la méthode des moments sont convergentes (loi faible des grands nombres) mais biaisées et généralement non efficaces.

c) Méthode des moments pondérés (MMP) (Greenwood et al., 1979)

La méthode des moments pondérés a été conçue pour les lois pouvant être exprimées explicitement sous forme inverse $X = X(F)$.

Soit un échantillon aléatoire x_1, x_2, \dots, x_n tiré d'une distribution $F(x; \theta_1, \theta_2, \dots, \theta_p)$. La méthode des moments pondérés conduit au système de p équations:

$$\beta_r = b_r \quad r = 1, 2, \dots, p \quad (2.16)$$

avec les moments pondérés de la distribution et de l'échantillon qui s'expriment respectivement par:

$$\beta_r = E(XF^r) = \int_0^1 x(F) F^r dF \quad (2.17a)$$

$$b_r = \sum_{i=1}^n \frac{(i-1)(i-2)\dots(i-r)}{(n-1)(n-2)\dots(n-r)} x_{(i)} \quad (2.17b)$$

où $x_{(i)}$ est le i ème élément de l'échantillon classé en ordre croissant.

d) Méthode généralisée des moments (Bobée et Ashkar, 1988)

Cette méthode est identique à la méthode des moments, exception faite qu'elle n'utilise pas nécessairement les p premiers moments mais des moments d'ordre r_1, r_2, r_3 . La méthode SAM (Sundry Averages Method), par exemple, utilise les moments non-centrés $r_1 = -1$ (moyenne harmonique), $r_2 = 0$ (moyenne géométrique), $r_3 = 1$ (moyenne arithmétique) et la méthode MM1 (Rao, 1980) les moments non-centrés $r_1 = 0, r_2 = 1, r_3 = 2$. À noter que le moment non-centré 0 est appelé moment d'ordre quasi-zéro.

2.3 DISTRIBUTIONS STATISTIQUES

2.3.1 Loi géométrique

Une variable aléatoire discrète X est distribuée selon une loi géométrique $GEO(q)$ si sa f.d.p est donnée par:

$$f(x) = q(1-q)^{x-1} \quad x = 1, 2, \dots \quad (2.18)$$

avec le paramètre $0 < q < 1$.

La moyenne et la variance de X sont données par $\mu'_1 = 1/q$ et $\mu_2 = (1-q)/q^2$.

a) Estimation du paramètre q

Soit un échantillon x_1, \dots, x_n . Le logarithme de la fonction de vraisemblance de la distribution géométrique est:

$$\ln L(q) = n \ln(q) + \left(\sum_{i=1}^n x_i - n \right) \ln(1-q) \quad (2.19)$$

L'estimateur du maximum de vraisemblance est obtenu en annulant la dérivée de $\ln L(q)$ par rapport à q , ce qui conduit à:

$$\hat{q} = 1/\bar{x} \quad (2.20)$$

La méthode des moments conduit exactement au même estimateur de q .

2.3.2 Loi "série logarithmique" (Johnson et Kotz, 1969)

Une variable aléatoire discrète X suit une distribution série logarithmique $SL(\theta)$ si sa f.d.p. est donnée par:

$$f(x) = \alpha \theta^x / x \quad x=1,2,\dots \quad (2.21a)$$

avec le paramètre $0 < \theta < 1$. La quantité α n'est pas un nouveau paramètre mais est donnée par

$$\alpha = -1/\ln(1-\theta) \quad (2.21b)$$

La moyenne et la variance de X sont respectivement:

$$\mu'_1 = \alpha \theta / (1-\theta) \quad (2.22a)$$

$$\mu_2 = \alpha \theta (1-\alpha \theta) / (1-\alpha)^2 \quad (2.22b)$$

a) Estimation du paramètre θ

Soit l'échantillon x_1, \dots, x_n . Le logarithme de la fonction de vraisemblance est

donné par:

$$\ln L(\theta) = n \ln(\alpha) + \sum_{i=1}^n x_i \ln(\theta) - \sum_{i=1}^n x_i \quad (2.23)$$

d'où on déduit l'estimateur de θ qui maximise $\ln L(\theta)$:

$$\bar{x} = -\frac{\hat{\theta}}{(1-\hat{\theta})\ln(1-\hat{\theta})} \quad (2.24)$$

L'équation (2.24) peut être résolue facilement par une méthode itérative. La méthode des moments est identique à la méthode du maximum de vraisemblance.

En effet, en utilisant l'équation (2.22a):

$$\begin{aligned} \mu'_1 &= \bar{x} \\ \hat{\alpha}\hat{\theta}/(1-\hat{\theta}) &= \bar{x} \\ -\frac{\hat{\theta}}{(1-\hat{\theta})\ln(1-\hat{\theta})} &= \bar{x} \end{aligned} \quad (2.25)$$

2.3.3 Loi binomiale négative tronquée (Johnson et Kotz, 1969)

La distribution binomiale négative tronquée est déduite d'une distribution binomiale négative BN à laquelle on a retranché la classe des valeurs nulles. Ainsi, la variable aléatoire discrète X suit une distribution binomiale négative tronquée BNT(r, b) si sa f.d.p est:

$$f(x) = \binom{x+r-1}{x} \frac{b^r(1-b)^x}{(1-b^r)} \quad x=1,2,\dots \quad (2.26a)$$

avec $0 < b < 1$, $r > 0$ et:

$$\binom{x+r-1}{x} = \frac{\Gamma(x+r)}{x!\Gamma(r)} \quad (2.26b)$$

où $\Gamma(\cdot)$ est la fonction gamma définie par:

$$\Gamma(\gamma) = \int_0^{\infty} u^{\gamma-1} e^{-u} du = (\gamma-1)\Gamma(\gamma-1) \quad \gamma > 0 \quad (2.26c)$$

Les distributions géométrique et série logarithmique sont des cas particuliers de la distribution binomiale négative tronquée. En effet, lorsque $r=1$ on obtient la distribution géométrique $GEO(q=b)$ et lorsque $r \rightarrow 0$ on a la distribution $LS(\theta=1-b)$ (voir l'annexe B).

Les moments non-centrés de la loi $BNT(r,b)$ peuvent être calculés en divisant les moments non-centrés de la loi $BN(r,b)$ par la quantité $(1-b')$ (Johnson et Kotz, 1969). La moyenne et la variance de X sont données par:

$$\mu'_1 = \frac{r(1-b)}{b(1-b')} \quad (2.27a)$$

$$\mu'_2 = \frac{r(1-b)}{b^2(1-b')} \left[1 - \frac{rb'(1-b)}{(1-b')} \right] \quad (2.27b)$$

a) Estimation des paramètres par la méthode des moments

En égalant les équations (2.27a) et (2.27b) aux moments respectifs de l'échantillon et en utilisant la relation:

$$f(x=1) = f_1 = rb'(1-b)/(1-b') \quad (2.28)$$

la méthode des moments, après quelques manipulations algébriques, conduit aux estimateurs:

$$\hat{b} = \bar{x}(1-\hat{f}_1)/s^2 \quad (2.29a)$$

$$\hat{r} = (\bar{x}\hat{b}-\hat{f}_1)/(1-\hat{b}) \quad (2.29b)$$

où \hat{f}_1 est la fréquence des valeurs 1 observées dans l'échantillon.

b) Estimation des paramètres par la méthode du maximum de vraisemblance

Soit un échantillon x_1, \dots, x_n . Le logarithme de la fonction de vraisemblance est donné par:

$$\ln L(r,b) = n[r \ln(b) - \ln(1-b)] + \ln(1-b) \sum_{i=1}^n x_i + \sum_{i=1}^n \sum_{j=1}^{x_i} \ln \left(\frac{r+j-1}{x_i-j+1} \right) \quad (2.30)$$

Les dérivées partielles par rapport aux paramètres de $\ln L(r,b)$ sont:

$$\frac{\partial \ln L}{\partial b} = \frac{nr}{b(1-b)} - \frac{1}{(1-b)} \sum_{i=1}^n x_i = 0 \quad (2.31a)$$

$$\frac{\partial \ln L}{\partial r} = \frac{n \ln(b)}{1-b} + \sum_{i=1}^n \sum_{j=1}^{x_i} \frac{1}{(r+j-1)} = 0 \quad (2.31b)$$

Le système d'équations (2.31) ne se résout pas aisément. Brass(1958) propose d'utiliser la relation (2.28) pour simplifier les équations de la méthode du maximum de vraisemblance. Après quelques calculs nous obtenons le système suivant:

$$b = \frac{(r+f_1)}{(r+\bar{x})} \quad (2.32a)$$

$$\frac{n\bar{x}(r+f_1)}{r(\bar{x}-f_1)} \ln(b) + \sum_{i=1}^n \sum_{j=1}^{x_i} \frac{1}{(r+j-1)} = 0 \quad (2.32b)$$

Le système d'équations (2.32) peut être résolu à l'aide d'une méthode itérative.

2.3.4 Loi normale

La distribution normale est certainement la loi statistique la plus employée et la plus connue dans tous les domaines scientifiques. La variable aléatoire continue X est distribuée selon une loi normale $N(\mu, \sigma^2)$ si sa f.d.p. est:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp \left(\frac{-[x-\mu]^2}{2\sigma^2} \right) \quad -\infty < x < \infty \quad (2.33)$$

La moyenne et la variance sont données respectivement par μ et σ^2 . De plus,

le coefficient d'asymétrie est nul et le coefficient d'aplatissement est égal à 3.

Les estimateurs de la méthode de vraisemblance sont déduits en maximisant le logarithme de la fonction de vraisemblance suivant:

$$\ln L(\mu, \sigma^2) = -\frac{n}{2} \ln(2\pi \sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n [x_i - \mu]^2 \quad (2.34a)$$

et ils sont donnés par:

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^n x_i = \bar{x} \quad (2.34b)$$

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n [x_i - \hat{\mu}]^2 \quad (2.34c)$$

En pratique, on utilise la variance non-biaisée $s^2 = n/(n-1) \hat{\sigma}^2$. Perreault et Bobée (1992a) donnent plus de détails sur les propriétés de la loi normale et sur l'estimation de ses paramètres.

2.3.5 Loi log-normale (2 paramètres)

Si la variable aléatoire $Y = \ln(X)$ suit une distribution normale $N(\mu_y, \sigma_y^2)$ alors X est distribuée selon une loi log-normale $LN(\mu_y, \sigma_y^2)$ et sa f.d.p. est donnée par:

$$f(x) = \frac{1}{x\sigma_y\sqrt{2\pi}} \exp\left(-\frac{[\ln(x) - \mu_y]^2}{2\sigma_y^2}\right) \quad x > 0 \quad (2.35)$$

Le moment non-centré d'ordre r est donné par:

$$\mu'_r = \exp(r\mu_y + r^2\sigma_y^2/2) \quad (2.36)$$

Ainsi, la moyenne, la variance et le coefficient d'asymétrie sont déduits de l'équation (2.36):

$$\mu'_1 = \exp(\mu_y + \sigma_y^2/2) \quad (2.37a)$$

$$\mu'_2 = \exp(2\mu_y + \sigma_y^2)(\exp(\sigma_y^2) - 1) = \mu_1'^2 \eta^2 \quad (2.37b)$$

$$C_s = \eta^3 + 3\eta \quad (2.37c)$$

où $\eta^2 = \exp(\sigma_y^2 - 1) - 1 = C_v^2$

a) Estimation des paramètres par la méthode des moments

En égalisant la moyenne et la variance de la distribution (équations (2.37a) et (2.37b)) avec la moyenne et la variance de l'échantillon et après quelques manipulations mathématiques, on obtient les estimateurs des paramètres:

$$\hat{\mu}_y = \ln(\bar{x} / [1 + s^2 / \bar{x}^2]^{1/2}) \quad (2.38a)$$

$$\hat{\sigma}_y^2 = \ln(1 + s^2 / \bar{x}^2) \quad (2.38b)$$

b) Estimation des paramètres par la méthode du maximum de vraisemblance

Soit un échantillon x_1, \dots, x_n . Le logarithme de la fonction de vraisemblance est donné par:

$$\ln L(\mu_y, \sigma_y^2) = -\frac{n}{2} \ln(2\pi \sigma_y^2) - \sum_{i=1}^n \ln(x_i) - \frac{1}{2\sigma_y^2} \sum_{i=1}^n [\ln(x_i) - \mu_y]^2 \quad (2.39)$$

Le maximum de $\ln L(\mu_y, \sigma_y^2)$ est solution des équations:

$$\frac{\partial \ln L}{\partial \mu_y} = \frac{1}{\sigma_y^2} \sum_{i=1}^n [\ln(x_i) - \mu_y] = 0 \quad (2.40a)$$

$$\frac{\partial \ln L}{\partial \sigma_y} = -\frac{n}{\sigma_y} + \frac{1}{\sigma_y^3} \sum_{i=1}^n [\ln(x_i) - \mu_y]^2 = 0 \quad (2.40b)$$

La résolution du système d'équations (2.40) donne les estimateurs:

$$\hat{\mu}_y = \frac{1}{n} \sum_{i=1}^n \ln(x_i) \quad (2.41a)$$

$$\hat{\sigma}_y^2 = \frac{1}{(n-1)} \sum_{i=1}^n [\ln(x_i) - \mu_y]^2 \quad (2.41b)$$

L'estimateur de σ_y^2 est multiplié par la quantité $n/(n-1)$ pour tenir compte du biais. On peut noter que la méthode du maximum de vraisemblance pour la loi log-

normale est équivalente à la méthode du maximum de vraisemblance de la loi normale appliquée à l'échantillon transformé par la fonction logarithmique.

2.3.6 Log-normale à 3 paramètres

Si la variable aléatoire $Y = \ln(X-m)$ suit une distribution normale $N(\mu_y, \sigma_y^2)$ alors X est distribuée selon une loi log-normale à trois paramètres $LN3(\mu_y, \sigma_y^2, m)$ et sa f.d.p. est donnée par:

$$f(x) = \frac{1}{(x-m)\sigma_y\sqrt{2\pi}} \exp\left(-\frac{[\ln(x-m)-\mu_y]^2}{2\sigma_y^2}\right) \quad x > m \quad (2.42)$$

Lorsque $m=0$, la variable X suit une distribution log-normale à 2 paramètres. Les moments de la variable X sont identiques à ceux de la loi log-normale à deux paramètres sauf en ce qui concerne la moyenne qui devient:

$$\mu'_1 = m + \exp(\mu_y + \sigma_y^2/2) \quad (2.43)$$

a) Estimation des paramètres par la méthode des moments

La méthode usuelle des moments consiste à égaler les équations (2.43), (2.37b) et (2.37c) respectivement avec la moyenne (\bar{x}), la variance (s^2) et le coefficient d'asymétrie (C_s) de l'échantillon. L'équation (2.37c) devient:

$$C_s = (\phi + 2)(\phi - 1)^{1/2} \quad (2.44)$$

où $\phi = \exp(\sigma_y^2)$. L'équation (2.44) ne dépend que de ϕ et la racine réelle s'écrit explicitement par (Spielge, 1989):

$$\hat{\phi} = [1 + (C_s^2 + \omega)/2]^{1/3} + [1 + (C_s^2 - \omega)/2]^{1/3} - 1 \quad (2.45)$$

où $\omega = C_s(4 + C_s^2)^{1/2}$. Ainsi, on déduit les estimateurs:

$$\hat{\mu}_y = s / [\hat{\phi}(\hat{\phi} - 1)]^{1/2} \quad (2.46a)$$

$$\hat{\sigma}_y^2 = \ln(\hat{\phi}) \quad (2.46b)$$

$$\hat{m} = \bar{x} - s / (\hat{\phi} - 1)^{1/2} \quad (2.46c)$$

Afin de corriger le biais du coefficient d'asymétrie pour les petits échantillons, la valeur de \hat{C}_s est modifiée pour donner (Bobée et Ashkar, 1991):

$$\hat{C}_s^* = \hat{C}_s (1 + 8.5/n) \quad (2.47)$$

b) Estimation des paramètres par la méthode du maximum de vraisemblance

Soit un échantillon x_1, \dots, x_n . Le logarithme de la fonction de vraisemblance est:

$$\ln L(\mu_y, \sigma_y, m) = -\frac{n}{2} \ln(2\pi \sigma_y^2) - \sum_{i=1}^n \ln(x_i - m) - \frac{1}{2\sigma_y^2} \sum_{i=1}^n [\ln(x_i - m) - \mu_y]^2 \quad (2.48)$$

Hill (1963) a montré que la fonction de vraisemblance $L(\mu_y, \sigma_y^2, m)$ devient très grand lorsque m tend vers $\min(x_i)$ et que le triplet $(-\infty, \infty, \min(x_i))$ est l'estimateur du maximum de vraisemblance. Malheureusement, ce triplet n'est pas applicable en pratique.

Il est toutefois possible d'obtenir des estimateurs qui conduisent à un maximum local de la fonction $L(\mu_y, \sigma_y^2, m)$. Il suffit d'annuler les premières dérivés de l'équation (2.48):

$$\frac{\partial \ln L}{\partial \mu_y} = \frac{1}{\sigma_y^2} \sum_{i=1}^n [\ln(x_i - m) - \mu_y] = 0 \quad (2.49a)$$

$$\frac{\partial \ln L}{\partial \sigma_y} = -\frac{n}{\sigma_y} + \frac{1}{\sigma_y^3} \sum_{i=1}^n [\ln(x_i - m) - \mu_y]^2 = 0 \quad (2.49b)$$

$$\frac{\partial \ln L}{\partial m} = \frac{1}{\sigma_y^2} \sum_{i=1}^n \frac{[\ln(x_i - m) - \mu_y]}{x_i - m} + \sum_{i=1}^n \frac{1}{(x_i - m)} = 0 \quad (2.49c)$$

En manipulant le système d'équations (2.49) on arrive aux équations:

$$\mu_y = \frac{1}{n} \sum_{i=1}^n \ln(x_i - m) \quad (2.50a)$$

$$\sigma_y^2 = \frac{1}{n} \sum_{i=1}^n [\ln(x_i - m)]^2 - \mu_y^2 \quad (2.50b)$$

$$(\sigma_y^2 - \mu_y) \left(\sum_{i=1}^n (x_i - m) \right)^{-1} + \sum_{i=1}^n \frac{1}{(x_i - m)} = 0 \quad (2.50c)$$

En substituant les équations (2.50a) et (2.50b) dans l'équation (2.50c) on obtient une expression qui ne dépend que du paramètre m . Cette équation peut être résolue par une méthode itérative afin de déterminer l'estimateur \hat{m} de m . Ensuite, on substitue \hat{m} dans les équations (2.50a) et (2.50b) pour obtenir les estimateurs de μ_y et σ_y^2 .

2.3.7 Loi Weibull

La f.d.p. d'une variable aléatoire X distribuée selon la loi Weibull $WE(\alpha, c)$ est:

$$f(x) = \frac{c}{\alpha} \left(\frac{x}{\alpha} \right)^{c-1} \exp[-(x/\alpha)^c] \quad x > 0 \quad (2.51)$$

avec les paramètres $c > 0$ et $\alpha > 0$. Lorsque $c = 1$, la distribution Weibull devient une distribution exponentielle $EX(1/\alpha)$.

La fonction de distribution est donnée sous forme explicite par:

$$F(x) = 1 - \exp[-(x/\alpha)^c] \quad (2.52)$$

La moyenne et la variance de X sont telles que:

$$\mu'_1 = \alpha \Gamma(1 + 1/c) \quad (2.53a)$$

$$\mu_2 = \alpha^2 [\Gamma(1 + 2/c) - \Gamma(1 + 1/c)^2] \quad (2.53b)$$

La méthode des moments revient à égaliser les équations (2.53a) et (2.53b)

respectivement à la moyenne et à la variance de l'échantillon.

Les estimateurs du maximum de vraisemblance sont obtenus en maximisant la fonction:

$$\ln L(\alpha, c) = n \ln(c) - n \ln(\alpha) - \sum_{i=1}^n (x_i/c)^c + (c-1) \sum_{i=1}^n (x_i/\alpha) \quad (2.54)$$

Ces deux méthodes d'estimation des paramètres de la loi Weibull sont présentées en détail par Perreault et Bobée (1992c).

2.3.8 Loi Gumbel

La variable aléatoire X est distribuée selon une loi Gumbel $GU(\alpha, u)$ si sa f.d.p. est donnée par:

$$f(x) = \frac{1}{\alpha} \exp\left[-\frac{(x-u)}{\alpha} - \exp\left(-\frac{(x-u)}{\alpha}\right)\right] \quad -\infty < x < \infty \quad (2.55)$$

avec les paramètres $\alpha > 0$ et $u > 0$.

La forme explicite de la fonction de distribution est donnée par:

$$F(x) = \exp[-\exp(-(x-u)/\alpha)] \quad (2.56)$$

La moyenne et la variance sont telles que:

$$\mu'_1 = u + C\alpha \quad (2.57a)$$

$$\mu_2 = \alpha^2 \pi^2 / 6 \quad (2.57b)$$

où $C=0.5772$ est la constante d'Euler. Pour cette distribution on a $C_s = 1.14$ et $C_k = 2.4$.

La méthode des moments est basée sur les équations (2.57a) et (2.57b) et la méthode des moments pondérés sur les moments pondérés 0 et 1 (Perreault et Bobée, 1992b). Les estimateurs du maximum de vraisemblance sont obtenus en maximisant le logarithme de la fonction de vraisemblance:

$$\ln L(\alpha, u) = -n \ln(\alpha) - \sum_{i=1}^n (x_i - u)/\alpha - \sum_{i=1}^n \exp[-(x_i - u)/\alpha] \quad (2.58)$$

Perreault et Bobée (1992b) donnent plus d'information sur ces trois méthodes d'estimation pour la loi Gumbel.

2.3.9 Loi généralisée des valeurs extrêmes (GEV)

Jenkinson (1955) a proposé la loi généralisée des valeurs extrêmes pour combiner les trois types de lois des valeurs extrêmes EV1 (GU), EV2 et EV3 développées par Fisher et Tippett (1928). La f.d.p. de la loi $GEV(\alpha, u, k)$ est telle que:

$$f(x) = \frac{1}{\alpha} \left[1 - \frac{k}{\alpha} (x-u) \right]^{1/k-1} \exp \left[- \left(1 - \frac{k}{\alpha} (x-u) \right)^{1/k} \right] \begin{cases} -\infty < x < u + \alpha/k & \text{si } k > 0 \\ u + \alpha/k < x < \infty & \text{si } k < 0 \end{cases} \quad (2.59)$$

avec les paramètres $\alpha > 0$, $u > 0$ et k .

La forme explicite de la fonction de distribution est donnée par:

$$F(x) = \exp[-(1 - k(x-u)/\alpha)^{1/k}] \quad (2.60)$$

Les distributions EV2 et EV3 sont déduites de la distribution GEV lorsque respectivement $k < 0$ et $k > 0$. Lorsque $k \rightarrow 0$, alors la distribution GEV tend vers la distribution Gumbel.

La forme générale des moments non-centrés d'ordre $r > 0$ pour les lois EV2 et EV3 est donnée par Perreault et Bobée (1992b). Ces moments n'existent que pour l'ordre r tel que:

$$r > -1/k \quad (2.61)$$

On obtient pour la moyenne, la variance et le coefficient d'asymétrie:

$$\mu'_1 = u + \alpha [1 - \Gamma(1+k)]/k \quad (2.62a)$$

$$\mu_2 = (\alpha/k)^2 [\Gamma(1+2k) - \Gamma(1+k)^2] \quad (2.62b)$$

$$C_s = -\frac{|k|}{k} \frac{[\Gamma(1+3k) - 3\Gamma(1+2k)\Gamma(1+k) + 2\Gamma(1+k)^3]}{[\Gamma(1+2k) - \Gamma(1+k)^2]^{3/2}} \quad (2.62c)$$

La méthode des moments est basée sur le système d'équations (2.62) et la méthode des moments pondérés sur les moments pondérés 0,1 et 2 (Perreault et Bobée, 1992b). La méthode du maximum de vraisemblance consiste à maximiser l'expression:

$$\ln L(\alpha, u) = -n \ln(\alpha) + (1/k - 1) \sum_{i=1}^n \ln[1 - k(x_i - u)/\alpha] - \sum_{i=1}^n [1 - k(x_i - u)/\alpha]^{1/k} \quad (2.63)$$

Perreault et Bobée (1992b) donnent plus d'information sur ces trois méthodes d'estimation de la loi GEV.

2.3.10 Loi gamma

La variable aléatoire X suit une distribution gamma $GA(\alpha, \lambda)$ si sa f.d.p. est:

$$f(x) = \frac{|\alpha|}{\Gamma(\lambda)} \exp(-\alpha x) (\alpha x)^{\lambda-1} \quad \begin{array}{ll} 0 \leq x < \infty & \text{si } \alpha > 0 \\ -\infty < x \leq 0 & \text{si } \alpha < 0 \end{array} \quad (2.64)$$

avec les paramètres $\alpha \neq 0$ et $\lambda > 0$. À remarquer que la distribution exponentielle est un cas particulier de la distribution gamma lorsque $\lambda = 1$.

La moyenne, la variance et le coefficient d'asymétrie de X sont données par:

$$\mu'_1 = \lambda/\alpha \quad (2.65a)$$

$$\mu_2 = \lambda/\alpha^2 \quad (2.65b)$$

$$C_s = \frac{\alpha}{|\alpha|} \frac{2}{\lambda^{1/2}} \quad (2.65c)$$

La méthode des moments consiste à égaler la moyenne et la variance théoriques avec la moyenne et la variance de l'échantillon. Les estimateurs du maximum de vraisemblance maximisent la fonction:

$$\ln L(\alpha, \lambda) = n\lambda \ln(|\alpha|) - n \ln(\Gamma(\lambda)) + (\lambda - 1) \sum_{i=1}^n \ln(x_i) - \alpha \sum_{i=1}^n x_i \quad (2.66)$$

Bobée et Ashkar (1991) traitent en détail de la loi gamma.

2.3.11 Loi gamma généralisée

La variable aléatoire X suit une distribution gamma généralisée $GG(\alpha, \lambda, s)$ si sa f.d.p. est:

$$f(x) = \frac{|s\alpha|}{\Gamma(\lambda)} \exp[-(\alpha x)^s] (\alpha x)^{s\lambda-1} \quad \begin{array}{ll} 0 \leq x < \infty & \text{si } \alpha > 0 \\ -\infty < x \leq 0 & \text{si } \alpha < 0 \end{array} \quad (2.67)$$

avec les paramètres d'échelle $\alpha \neq 0$, $\lambda > 0$ et $s \neq 0$. La distribution gamma est un cas particulier de la distribution gamma généralisée obtenu lorsque $s=1$.

Le moment non-centré d'ordre r est donné par:

$$\mu'_r = \frac{1}{\alpha^r} \frac{\Gamma(\lambda + 1/s)}{\Gamma(\lambda)} \quad r/s > -\lambda \quad (2.68)$$

Le logarithme de la fonction de vraisemblance s'écrit:

$$\ln L(\alpha, \lambda, s) = n \ln(|s|) + n \lambda s \ln(\alpha) - n \ln(\Gamma(\lambda)) + (s\lambda - 1) \sum_{i=1}^n \ln(x_i) - \alpha^s \sum_{i=1}^n x_i^s \quad (2.69)$$

La méthode du maximum de vraisemblance, la méthode des moments et la méthode généralisée des moments (SAM et MM1) appliquées à la loi gamma généralisée sont décrites par Bobée et Ashkar (1991).

2.3.12 Loi Pearson type 3

La variable aléatoire X suit une distribution Pearson type 3 $P3(\alpha, \lambda, m)$ si sa f.d.p. s'écrit:

$$f(x) = \frac{|\alpha|}{\Gamma(\lambda)} \exp[-\alpha(x-m)] [\alpha(x-m)]^{\lambda-1} \quad \begin{array}{ll} m \leq x < \infty & \text{si } \alpha > 0 \\ -\infty < x \leq m & \text{si } \alpha < 0 \end{array} \quad (2.70)$$

avec les paramètres $\alpha \neq 0$, $\lambda > 0$ et m . La distribution gamma est un cas particulier

de la distribution Pearson type 3 obtenu lorsque $m=0$.

La variance et le coefficient d'asymétrie sont les mêmes que celles de la distribution gamma (relation 2.65b et 2.65c). La moyenne de X est telle que:

$$\mu'_1 = m + \lambda/\alpha \quad (2.71)$$

La méthode des moments considère la moyenne, la variance et le coefficient d'asymétrie. Le biais du coefficient d'asymétrie de la population est réduit à l'aide de l'équation (2.47). La méthode du maximum de vraisemblance consiste à maximiser l'expression:

$$\ln L(\alpha, \lambda, m) = +n\lambda \ln(|\alpha|) - n \ln(\Gamma(\lambda)) + (\lambda - 1) \sum_{i=1}^n \ln(x_i - m) - \alpha \sum_{i=1}^n (x_i - m) \quad (2.72)$$

La méthode du maximum de vraisemblance ne produit pas de solution lorsque $\lambda \geq 1$ ($C_s \leq 2$). Une façon de contourner ce problème est d'utiliser la méthode du maximum de vraisemblance conditionnel (Bobée et Ahskar, 1991). Ces deux méthodes d'estimation appliquées à la loi Pearson type 3 sont détaillées par Bobée et Ashkar (1991).

2.3.13 Loi log-Pearson type 3

Si la variable aléatoire $Y = \log_a(X)$ suit une distribution $P3(\alpha, \lambda, m)$, alors X est distribuée selon une loi log-Pearson type 3 $LP3(\alpha, \lambda, m)$ et sa f.d.p. s'écrit:

$$f(x) = \frac{k|\alpha|}{x\Gamma(\lambda)} \exp[-\alpha(\log_a(x) - m)] [\alpha(\log_a(x) - m)]^{\lambda-1} \begin{matrix} e^{m/k} \leq x < \infty & \text{si } \alpha > 0 \\ 0 \leq x \leq e^{m/k} & \text{si } \alpha < 0 \end{matrix} \quad (2.73)$$

avec les paramètres $\alpha \neq 0$, $\lambda > 0$ et m . De plus, $k = 1/\ln(a)$ où a est la base du logarithme.

Le moment non-centré d'ordre r est donné par:

$$\mu'_r = \exp(mr/k) / (1 - r/(\alpha k))^{\lambda} \quad r < \alpha k \quad \text{si } \alpha > 0 \quad (2.74)$$

Le logarithme de la fonction de vraisemblance est tel que:

$$\begin{aligned} \ln L(\alpha, \lambda, m) = & n \ln(\alpha |k) - n \ln(\Gamma(\lambda)) + n \alpha m - (1 + \alpha k) \sum_{i=1}^n \ln(x_i) \\ & - (\lambda - 1) \sum_{i=1}^n \ln[\alpha (k \ln(x_i) - m)] \end{aligned} \quad (2.75)$$

La méthode du maximum de vraisemblance appliquée à la loi log-Pearson type 3 est équivalente à la méthode du maximum de vraisemblance dans le cas de la loi Pearson type 3 sur l'échantillon transformé ($y = \log_a(x)$). L'estimation des paramètres de la distribution log-Pearson type 3 par la méthode du maximum de vraisemblance, la méthode des moments et la méthode généralisée des moments (SAM et MM1) est décrite par Bobée et Ashkar (1991).

2.3.14 Loi log-gamma

Si la variable aléatoire $Y = \log_a(X)$ suit une distribution $GA(\alpha, \lambda)$, alors X est distribuée selon une loi log-gamma $LG(\alpha, \lambda)$. La loi log-gamma est un cas particulier de la loi log-Pearson type 3 lorsque $m=0$ et nous référons donc à la section 2.3.13 pour plus de détails. Mentionnons, toutefois, que si la loi log-gamma est applicable avec un système d'unités de mesure (A) elle ne sera pas applicable si l'unité de mesure est changée (annexe C). Les méthodes d'estimation des paramètres considérées dans notre étude sont la méthode des moments sur l'échantillon et les méthodes des moments et du maximum de vraisemblance sur l'échantillon transformé.

2.3.15 Loi exponentielle

Une variable aléatoire X est distribuée selon une loi exponentielle $EX(\alpha)$ si sa f.d.p. est donnée par:

$$f(x) = \alpha \exp(-\alpha x) \quad x \geq 0 \quad (2.76)$$

avec le paramètre $\alpha > 0$.

La fonction de distribution est telle que:

$$F(x) = 1 - \exp(-\alpha x) \quad (2.77)$$

Le moment non-centré d'ordre r est:

$$\mu'_r = r!/\alpha^r \quad (2.78)$$

La moyenne et la variance de X sont déduites de (2.78) et s'expriment par $\mu'_1 = 1/\alpha$ et $\mu'_2 = 1/\alpha^2$. De plus, $C_s = 2$ et $C_k = 9$.

a) Estimation du paramètre α

Soit un échantillon aléatoire x_1, \dots, x_n . Le logarithme de la fonction de vraisemblance de la loi exponentielle est:

$$\ln L(\alpha) = n \ln(\alpha) - \alpha \sum_{i=1}^n x_i \quad (2.79)$$

L'estimateur du maximum de vraisemblance est déterminé en annulant la dérivée de $\ln L(\alpha)$ par rapport à α , ce qui conduit à:

$$\frac{\partial \ln L}{\partial \alpha} = \frac{n}{\alpha} - \sum_{i=1}^n x_i = 0 \quad (2.80a)$$

D'où:

$$\hat{\alpha} = 1/\bar{x} \quad (2.80b)$$

Dans le cas de la loi exponentielle, la méthode des moments est équivalente à la méthode du maximum de vraisemblance.

2.3.16 Loi exponentielle mixte

La distribution exponentielle mixte est composée de la somme de 2 distributions exponentielles $EX(\alpha_1 = 1/a_1)$ et $EX(\alpha_2 = 1/a_2)$. Ainsi, la variable aléatoire X est distribuée selon une distribution exponentielle mixte $EXM(a_1, a_2, p)$ si sa f.d.p. est:

$$f(x) = \frac{p}{a_1} \exp\left(-\frac{x}{a_1}\right) + \frac{(1-p)}{a_2} \exp\left(-\frac{x}{a_2}\right) \quad x \geq 0 \quad (2.81)$$

avec les paramètres $a_1 > a_2 > 0$ et $0 \leq p \leq 1$. Lorsque $p=0$ ou $p=1$, la distribution exponentielle mixte devient une simple distribution exponentielle.

La fonction distribution est donnée par:

$$F(x) = 1 - p \exp(-x/a_1) - (1-p) \exp(-x/a_2) \quad (2.82)$$

Le moment non-centré d'ordre r de X est déduit du moment non-centré de la loi exponentielle (2.78), ce qui donne:

$$\mu'_r = r! [p(a_1^r - a_2^r) + a_2^r] \quad (2.83)$$

Ainsi, la moyenne et la variance de X sont:

$$\mu'_1 = p(a_1 - a_2) + a_2 \quad (2.84a)$$

$$\mu'_2 = 2p(a_1 - a_2)[a_1 - p(a_1 - a_2)] + a_2^2 \quad (2.84b)$$

a) Estimation des paramètres par la méthode des moments

La méthode des moments consiste à évaluer les 3 premiers moments non-centrés (équation (2.83) avec $r = 1, 2, 3$) avec les moments équivalents de l'échantillon.

Considérons les quantités:

$$A = a_1 + a_2 \quad (2.85a)$$

$$B = a_1 a_2 \quad (2.85b)$$

Après quelques manipulations algébrique, on obtient des expressions de A et de B qui ne dépendent pas de p :

$$\hat{A} = \frac{(m'_3/6 - \bar{x}m'_2/2)}{(m'_2/2 - \bar{x}^2)} \quad (2.86a)$$

$$\hat{B} = m'_r \hat{A} - m'_2/2 \quad (2.86b)$$

Les estimateurs de a_1 et a_2 sont déterminés en solutionnant le système d'équation (2.85):

$$\hat{a}_1 = \frac{\hat{A} + (\hat{A}^2 - 4\hat{B})^{1/2}}{2} \quad (2.87a)$$

$$\hat{a}_2 = \frac{\hat{A} - (\hat{A}^2 - 4\hat{B})^{1/2}}{2} \quad (2.87b)$$

et le paramètre p est donné par:

$$\hat{p} = (\bar{x} - \hat{a}_2) / (\hat{a}_1 - \hat{a}_2) \quad (2.87c)$$

b) Estimation des paramètres par la méthode du maximum de vraisemblance

Soit un échantillon aléatoire x_1, \dots, x_n . Le logarithme de la fonction de vraisemblance est donné par:

$$\ln L(a_1, a_2, p) = \sum_{i=1}^n \ln \left[\frac{p}{a_1} \exp\left(-\frac{x_i}{a_1}\right) + \frac{(1-p)}{a_2} \exp\left(-\frac{x_i}{a_2}\right) \right] \quad (2.88)$$

Everitt et Hand (1981) donnent une méthode itérative pour résoudre les équations maximisant le logarithme de la fonction de vraisemblance:

$$\hat{p}_k = \frac{1}{n} \sum_{i=1}^n \frac{\hat{b}_{1,k}(x_i)}{\hat{b}_{1,k}(x_i) + \hat{b}_{2,k}(x_i)} \quad (2.89a)$$

$$\hat{a}_{1,k} = \frac{1}{n \hat{p}_k} \sum_{i=1}^n \frac{x_i \hat{b}_{1,k}(x_i)}{\hat{b}_{1,k}(x_i) + \hat{b}_{2,k}(x_i)} \quad (2.89b)$$

$$\hat{a}_{2,k} = \frac{1}{n(1 - \hat{p}_k)} \sum_{i=1}^n \frac{x_i \hat{b}_{2,k}(x_i)}{\hat{b}_{1,k}(x_i) + \hat{b}_{2,k}(x_i)} \quad (2.89c)$$

avec

$$\hat{b}_{1,k}(x_i) = \frac{\hat{p}_{k-1}}{\hat{a}_{1,k-1}} \exp\left(-\frac{x_i}{\hat{a}_{1,k-1}}\right) \quad (2.90a)$$

$$\hat{b}_{2,k}(x_i) = \frac{(1 - \hat{p}_{k-1})}{\hat{a}_{2,k-1}} \exp\left(-\frac{x_i}{\hat{a}_{2,k-1}}\right) \quad (2.90b)$$

Les paramètres estimés par la méthode des moments peuvent servir de valeurs initiales pour débiter l'itération.

2.3.17 Loi Kappa (2 paramètres)

La distribution Kappa est déduite de la famille des distributions bêta. Une variable aléatoire X est distribuée selon une loi Kappa à deux paramètres $KA(\alpha, \beta)$

si sa f.d.p. est:

$$f(x) = (\alpha/\beta)[\alpha + (x/\beta)^\alpha]^{-(1+1/\alpha)} \quad x > 0 \quad (2.91)$$

avec les paramètres $\alpha > 0$ et $\beta > 0$.

La fonction de distribution est donnée par:

$$F(x) = (x/\beta)[\alpha + (x/\beta)^\alpha]^{-1/\alpha} \quad (2.92)$$

Le moment non-centré d'ordre r s'écrit:

$$\mu'_r = \beta^r \alpha^{-w_r} B(z_r, w_r) \quad (2.93)$$

où $z_r = (1+r)/\alpha$ et $w_r = (\alpha-r)/\alpha$. Et la fonction bêta $B(.,.)$ est définie par:

$$B(m,n) = \int_0^1 t^{m-1} (1-t)^{n-1} dt = \Gamma(m)\Gamma(n)/\Gamma(m+n) \quad m,n > 0 \quad (2.94)$$

Le moment d'ordre r existe seulement si $w_r > 0$ (ou $\alpha > r$). La moyenne et la variance s'écrivent:

$$\mu'_1 = \beta \alpha^{-(\alpha-1)/\alpha} B(2/\alpha, (\alpha-1)/\alpha) \quad (2.95a)$$

$$\mu'_2 = \beta^2 \alpha^{-(\alpha-2)/\alpha} [B(3/\alpha, (\alpha-2)/\alpha) - \alpha^{-1} B(2/\alpha, (\alpha-1)/\alpha)^2] \quad (2.95b)$$

a) Estimation des paramètres par la méthode des moments

Les estimateurs des paramètres de la méthode des moments s'obtiennent en égalant les 2 premiers moments non-centrés de la population (équation (2.93) avec $r=1,2$) avec ceux de l'échantillon. En effectuant le quotient $\mu'_2/\mu_1'^2$, on obtient une expression qui ne dépend pas de β :

$$g(\alpha) = \frac{m'_2}{\bar{x}^2} = \frac{\alpha B(3/\alpha, (\alpha-2)/\alpha)}{B(2/\alpha, (\alpha-1)/\alpha)^2} \quad (2.96a)$$

L'équation (2.96a) se résout par une méthode numérique. Une fois la valeur de $\hat{\alpha}$ connue, l'estimateur de β est déterminé par:

$$\hat{\beta} = \frac{\bar{x} \hat{\alpha}^{(\hat{\alpha}-1)/\hat{\alpha}}}{B(2/\hat{\alpha}, (\hat{\alpha}-1)/\hat{\alpha})} \quad (2.96b)$$

b) Estimation des paramètres par la méthode du maximum de vraisemblance

Soit un échantillon aléatoire x_1, \dots, x_n . Le logarithme de la fonction de vraisemblance est donné par:

$$\ln L(\alpha, \beta) = n \ln(\alpha/\beta) - (1 + 1/\alpha) \sum_{i=1}^n \ln(\alpha + (x_i/\beta)^\alpha) \quad (2.97)$$

Les équations à résoudre pour maximiser la vraisemblance s'écrivent:

$$\frac{\partial \ln L}{\partial \alpha} = \frac{n}{\alpha} + \frac{1}{\alpha^2} \sum_{i=1}^n \ln[\alpha + (x_i/\beta)^\alpha] - (1 + 1/\alpha) \sum_{i=1}^n \frac{1 + (x_i/\beta)^\alpha \ln(x_i/\beta)}{\alpha + (x_i/\beta)^\alpha} = 0 \quad (2.98a)$$

$$\frac{\partial \ln L}{\partial \beta} = -\frac{n}{\beta} + \frac{(\alpha + 1)}{\beta} \sum_{i=1}^n \frac{(x_i/\beta)^\alpha}{\alpha + (x_i/\beta)^\alpha} = 0 \quad (2.98b)$$

Le système d'équation (2.98) peut être résolu à l'aide de la méthode de Newton-Raphson.

CHAPITRE 3 THÉORIE DE L'AJUSTEMENT

Ce chapitre a pour objet d'établir les bases théoriques de notre étude. Dans un premier temps, on examinera les hypothèses de base que doivent respecter les échantillons afin de pouvoir leur ajuster une distribution. Dans un deuxième temps, on introduira les tests appropriés pour vérifier ces hypothèses. Finalement, on discutera des méthodes permettant de mesurer la qualité des ajustements.

3.1 CRITÈRES STATISTIQUES

L'ajustement d'une distribution statistique nécessite la connaissance d'un échantillon de taille n constitué de réalisations indépendantes et identiquement distribuées de la variable aléatoire considérée (i.i.d.). Il est donc essentiel de vérifier l'indépendance de l'échantillon. De plus, il faut s'assurer de la stationnarité et de l'homogénéité de l'échantillon et examiner les valeurs singulières pour vérifier si l'échantillon est représentatif de la population étudiée.

3.1.1 Indépendance

Les événements A et B sont indépendants si l'occurrence ou la non-occurrence de l'un n'influence pas l'occurrence ou la non-occurrence de l'autre ($P[A|B]=P[A]$ et $P[B|A]=P[B]$). En termes mathématiques, les événements A et B sont indépendants si et seulement si:

$$P[A \text{ et } B] = P[A] \cdot P[B] \quad (3.1)$$

En ce qui concerne les précipitations, la dépendance varie en fonction de l'intervalle entre les éléments successifs de la série: le lien entre les hauteurs successives de pluie horaire est plutôt grand tandis que celui des maximums annuels est plutôt faible. Le test de Wald-Wolfowitz (1943) permet de vérifier l'indépendance de l'échantillon.

Le coefficient de corrélation donne une mesure de la tendance des points (z_i, y_i) à se concentrer autour de la droite de régression $E(Y|z) = az + b$ (Rice, 1987). Le coefficient de corrélation des points (z_i, y_i) est défini par:

$$\hat{\rho} = \frac{\sum_{i=1}^n (y_i - \bar{y})(z_i - \bar{z})}{\left[\sum_{i=1}^n (y_i - \bar{y})^2 \sum_{i=1}^n (z_i - \bar{z})^2 \right]^{1/2}} \quad (3.2)$$

La valeur de $\hat{\rho}$ qui estime le coefficient de corrélation ρ entre les deux variables Z et Y varie entre -1 et 1. Le lien linéaire entre les deux variables est fort lorsque ρ est près de 1 en valeur absolue. Lorsque $\rho = 0$, le lien entre les variables est nul.

Le coefficient d'autocorrélation d'ordre un $\hat{\rho}_1$ est mesurée entre les valeurs successives d'un échantillon. Afin de calculer $\hat{\rho}_1$ d'un échantillon X_i , on remplace z_i par x_i et y_i par x_{i+1} dans l'équation (3.2) et $i = 1, 2, \dots, n-1$.

3.1.2 Stationnarité

La stationnarité implique qu'en moyenne les paramètres statistiques (moyenne, variance, moment d'ordre 3, etc...) sont invariants avec le temps. Les sauts, les cycles et les tendances sont des types de non-stationnarités. La stationnarité des séries pluviométriques peut être affectée par l'urbanisation et les fluctuations climatiques à long terme. Le test de Mann-Whitney (1947) pour les sauts et le test de Wald-Wolfowitz (1943) pour les tendances rend possible la vérification de la stationnarité.

3.1.3 Homogénéité

La condition de représentativité implique que:

- l'échantillon est formé de réalisations de la même variable;
- les éléments de l'échantillon doivent provenir de la même population statistique.

Un changement dans l'emplacement du pluviomètre, la construction de nouveaux édifices ou la pousse des arbres près du pluviomètre ainsi qu'un remplacement de l'appareil de mesure peuvent altérer l'homogénéité de la série de données. De plus, la saisonnalité des précipitations peut entraîner une séparation de l'échantillon en différentes saisons. Le test de Mann-Whitney (1947) pour deux sous-échantillons et le test de Kruskal-Wallis (1952) pour plusieurs sous-échantillons permettent de vérifier l'homogénéité d'un échantillon.

3.1.4 Valeurs singulières

Il est important de vérifier si une valeur singulière appartient ou non à la même population que le reste de l'échantillon. Une valeur singulière peut provenir d'une pluie causée par une situation climatique très différente de celles des autres pluies de l'échantillon ou par une erreur de lecture. Il existe des valeurs singulières inférieures et supérieures. La présence d'une ou de plusieurs valeurs singulières peut affecter considérablement la qualité d'un ajustement. Le test de Grubbs-Beck (1972) permet de détecter les valeurs singulières dans un échantillon.

3.2 TESTS STATISTIQUES

Un test statistique permet de rejeter ou de ne pas rejeter une hypothèse nulle (H_0) face à une hypothèse alternative (H_1). La décision de rejeter ou non H_0 est faite sur la base de T , où T est une statistique obtenue à partir de l'échantillon. Lorsque la fonction de distribution de T est connue, il est possible de définir une région de rejet $\{T > t\}$ qui a la probabilité α sous l'hypothèse nulle. La statistique T est alors comparée à la valeur critique t qui sépare la région de rejet et la région de non-rejet. Lorsque l'hypothèse H_0 n'est pas rejetée, cela signifie que rien dans les observations ne permet de rejeter l'hypothèse H_0 qui est à la base du test statistique employé.

Deux types d'erreur peuvent se produire en appliquant un test statistique:
Erreur de type I : L'hypothèse H_0 est rejetée alors qu'elle est vraie. La probabilité de commettre une erreur de type I est représentée par α .

Erreur de type II : L'hypothèse H_0 est acceptée alors qu'elle est fautive. La probabilité de commettre une erreur de type II est représentée par β .

La probabilité α est appelée le niveau de signification du test et elle est fixée d'avance à une petite valeur (0.05 ou 0.01). La puissance du test est égale à $1-\beta$.

3.2.1 Test de Wald-Wolfowitz (1943)

Le test de Wald-Wolfowitz permet de vérifier l'indépendance et de détecter les tendances dans un échantillon. Ce test est non-paramétrique, c'est-à-dire qu'il ne suppose pas que l'échantillon provienne d'une loi particulière (par exemple, la loi normale).

Soit un échantillon x_1, \dots, x_n . On considère la statistique R telle que:

$$R = \sum_{i=1}^{n-1} x_i x_{i+1} + x_1 x_n \quad (3.3)$$

Dans le cas où les éléments de l'échantillon sont indépendants, R suit approximativement une distribution normale de moyenne et de variance:

$$\bar{R} = \frac{S_1^2 - S_2}{n-1} \quad (3.4)$$

$$Var(R) = \frac{S_2^2 - S_4}{n-1} + \frac{S_1^2 - 4S_1^2 S_2 + 4S_1 S_3 + S_2^2 - 2S_4}{(n-1)(n-2)} - \bar{R}^2 \quad (3.5)$$

où $S_r = N m_r'$ et m_r' est le moment non-centré d'ordre r de l'échantillon défini à l'équation (2.6a).

La quantité $U = (R - \bar{R}) / \sqrt{Var(R)}$ suit approximativement une distribution normale centrée réduite et peut être utilisée pour tester l'indépendance d'un échantillon. L'hypothèse d'indépendance n'est pas rejetée au niveau de signification

α si $|u| < u_{\alpha/2}$, où $u_{\alpha/2}$ est la variable normale dont la probabilité au dépassement est $\alpha/2$.

En pratique, on utilise $\alpha = 5\%$ et $\alpha = 1\%$:

Si $|u| < 1.96$: l'hypothèse d'indépendance n'est pas rejetée à un niveau de signification de 5%. Il n'y a pas de lien entre les éléments de l'échantillon.

Si $1.96 < |u| < 2.57$: l'hypothèse d'indépendance n'est pas rejetée à un niveau de signification de 1% mais elle est rejetée à un niveau de signification de 5%. L'indépendance est douteuse.

Si $|u| > 2.57$: l'hypothèse d'indépendance est rejetée à un niveau de signification de 1%. Il y a donc un lien entre les valeurs successives de l'échantillon.

3.2.2 Test de Mann-Whitney (1947)

Ce test non-paramétrique permet de vérifier l'homogénéité d'un échantillon. Il est analogue au test paramétrique de comparaison des moyennes t.

Considérons deux sous-échantillons de taille p et q (où $p \leq q$) regroupés en un échantillon total de taille $n = p + q$ classé en ordre croissant. Soient les quantités V et W tel que:

$$V = R - \frac{p(p+1)}{2} \quad (3.6a)$$

$$W = pq - V \quad (3.6b)$$

où R : est la somme des rangs des éléments du plus petit sous-échantillon (de taille p) dans l'échantillon total;

V : représente le nombre de dépassement des éléments du deuxième sous-échantillon par ceux du premier sous-échantillon;

W : représente le nombre de dépassement des éléments du premier sous-

échantillon par ceux du deuxième sous-échantillon.

La statistique U est définie par la plus petite valeur entre V et W . Lorsque $N > 20$, et $p, q > 3$ et si les deux sous-échantillons proviennent d'une même population, U est approximativement distribuée normalement avec une moyenne et une variance:

$$\bar{U} = \frac{pq}{2} \quad (3.7)$$

$$\text{Var}(U) = pq(n+1)/12 \quad (3.8)$$

La quantité $u = (U - \bar{U}) / \sqrt{\text{Var}(U)}$ suit approximativement une distribution normale centrée réduite et peut être utilisée pour tester l'homogénéité d'un échantillon. L'hypothèse d'homogénéité n'est pas rejetée au niveau de signification α si $|u| < u_{\alpha/2}$, où $u_{\alpha/2}$ est la variable normale dont la probabilité au dépassement est $\alpha/2$.

En pratique, on utilise $\alpha = 5\%$ et $\alpha = 1\%$:

- Si $|u| < 1.96$: l'hypothèse d'homogénéité n'est pas rejetée à un niveau de signification de 5%. Les deux sous-échantillons proviennent de la même population.
- Si $1.96 < |u| < 2.57$: l'hypothèse d'homogénéité n'est pas rejetée à un niveau de signification de 1% mais elle est rejetée à un niveau de signification de 5%. L'homogénéité est douteuse.
- Si $|u| > 2.57$: l'hypothèse d'homogénéité est rejetée à un niveau de signification de 1%. Les deux sous-échantillons proviennent donc de deux populations différentes.

Dans le cas où la taille du plus grand sous-échantillon est trop faible pour obtenir une approximation par la loi normale, on utilise la statistique R^* :

$$R^* = \min(R, R') \quad (3.9)$$

où R est définie comme précédemment et $R' = p(n+1)-R$. L'hypothèse d'homogénéité des deux sous-échantillons n'est pas rejetée à un niveau de signification α si $R^* > R^*_{\alpha}$, où R^*_{α} est la valeur critique de R^* (Rice, 1987).

3.2.3 Test de Kruskal-Wallis (1952)

Le test de Kruskal-Wallis est une généralisation du test de Mann-Whitney. Il est également non-paramétrique et permet de déterminer si plusieurs sous-échantillons proviennent d'une même population. Ce test est utile pour détecter les variations saisonnières.

Soient k sous-échantillons de taille n_1, \dots, n_k regroupés en un échantillon total de taille $n = n_1 + \dots + n_k$ classé en ordre croissant. On considère la statistique:

$$\tau = \frac{12}{n(n+1)} \sum_{j=1}^k \frac{R_j^2}{n_j} - 3(n+1) \quad (3.10)$$

où R_j est la somme des rangs du sous-échantillon j dans l'échantillon total.

Lorsque $n_j > 5$, τ suit approximativement une distribution chi-2 avec $k-1$ degrés de liberté. L'hypothèse d'homogénéité des sous-échantillons n'est pas rejetée à un niveau de signification α si $\tau < \chi^2_{k-1, \alpha}$, où $\chi^2_{k-1, \alpha}$ est la variable chi-2 à $k-1$ degrés de liberté dont la probabilité au dépassement est α . Dans le cas contraire, il y a au moins un sous-échantillon qui ne proviennent pas de la même population.

Dans cette étude, on s'intéresse à $k=4$. De plus, en pratique, on utilise $\alpha = 5\%$ et $\alpha = 1\%$:

Si $|u| < 7.81$: l'hypothèse d'homogénéité n'est pas rejetée à un niveau de signification de 5%. Les 4 sous-échantillons proviennent de la même population.

Si $7.81 < |u| < 11.34$: l'hypothèse d'homogénéité n'est pas rejetée à un niveau

de signification de 1% mais elle est rejetée à un niveau de signification de 5%. L'homogénéité est douteuse.

Si $|u| > 11.34$

: l'hypothèse d'homogénéité est rejetée à un niveau de signification de 1%. Un des 4 sous-échantillons provient donc d'une population différente.

3.2.4 Test de Grubbs-Beck (1972)

Le test de Grubbs-Beck permet de vérifier si les valeurs extrêmes appartiennent à la même population que le reste de l'échantillon. Ce test suppose que les données proviennent d'une population normale. Pour un échantillon ayant une asymétrie positive on fait l'hypothèse que les données sont distribuées selon une loi log-normale. Puisque ce test est peu robuste, il ne sera utilisé qu'à titre indicatif (erreur de lecture). Les observations détectées seront ensuite validées physiquement à l'aide des stations avoisinantes.

Soient les deux quantités:

$$X_s = \bar{x} + K_N s \quad (3.11a)$$

$$X_l = \bar{x} - K_N s \quad (3.11b)$$

où K_N est la statistique de Grubbs-Beck tabulée pour différentes tailles d'échantillon et pour différents niveaux de signification. \bar{x} et s^2 sont respectivement la moyenne et la variance non-biaisée de l'échantillon.

Les valeurs supérieures à X_s sont considérées des valeurs singulières supérieures et les valeurs inférieures à X_l sont considérées des valeurs singulières inférieures.

3.3 CRITÈRES D'ADÉQUATION

Plusieurs tests peuvent être utilisés pour examiner si la distribution choisie rend bien compte des observations. Par exemple, les tests du chi-2 et de Kolmogorov-Smirnov sont communément employés. Ces deux tests ont

l'inconvénient de supposer les paramètres de la population connus à priori et non pas estimés à partir d'un échantillon. Dans les paragraphes qui suivent nous examinons plus en détail le test du chi-2, le test du rapport des vraisemblances maximales et le critère d'Akaike. Le test de Kolmogorov-Smirnov n'est pas considéré ici.

3.3.1 Test du chi-2

Nous résumons ici le concept de base du test du chi-2. L'intervalle total de variation des observations est divisé en k classes mutuellement exclusives et exhaustives. Chacune des classes ($j=1, \dots, k$) a un nombre de valeurs observés O_j et un nombre de valeurs théoriques correspondantes E_j . On peut montrer que:

$$\chi_c^2 = \sum_{j=1}^k \frac{(O_j - E_j)^2}{E_j} \quad (3.12)$$

est asymptotiquement distribué selon une loi chi-2 avec $\nu = k - p - 1$ degrés de liberté lorsque les p paramètres ont été estimés à partir de l'échantillon.

La validité de l'ajustement n'est pas rejetée à un niveau de signification α si $\chi_c^2 < \chi_{\nu, \alpha}^2$, où $\chi_{\nu, \alpha}^2$ est la variable chi-2 à ν degrés de liberté dont la probabilité au dépassement est α .

La p-value est définie par:

$$p = P[\chi_{\nu}^2 \geq \chi_c^2] \quad (3.13)$$

L'interprétation de la p-value se fait de la manière suivante:

- $p > 0.05$: l'adéquation n'est pas rejetée à un niveau de signification de 5%;
- $0.01 < p < 0.05$: l'adéquation est rejetée à un niveau de signification de 5% mais elle ne l'est pas à 1%;
- $p < 0.01$: l'adéquation est rejetée à un niveau de signification de 1%.

Il n'existe pas de règle universelle pour déterminer le nombre de classes. Cependant, en pratique, on choisit un nombre de classe de sorte que les fréquences espérées dans chaque classe soient supérieures ou égales à 5. Dans le cas de grands échantillons, il n'est pas nécessaire d'avoir plus de 20 classes car la puissance du test n'est pas augmentée au-delà de cette valeur.

Pour de petits échantillons comme ceux rencontrés en hydrométéorologie, le test du chi-2 n'est pas suffisamment sensible pour discriminer les distributions et les méthodes d'estimation entre-elles.

3.3.2 Le test du rapport des vraisemblances maximales

Le test du rapport des vraisemblances maximales permet de comparer deux distributions de la même famille dont les paramètres ont été estimés par la méthode du maximum de vraisemblance.

Soient les hypothèses:

H_0 : les observations proviennent de la loi F_0 (p paramètres)

H_1 : les observations proviennent de la loi F_1 ($(p+k)$ paramètres)

On définit la statistique:

$$t = 2(\ln L[H_1] - \ln L[H_0]) \quad (3.14)$$

où $L[H_0]$ et $L[H_1]$ sont respectivement les fonctions de vraisemblance de la loi F_0 et F_1 . La loi F_0 appartient à une sous-famille de la loi F_1 . On peut montrer que la statistique t est asymptotiquement distribuée selon une loi chi-2. Le nombre de degré de liberté est égal au nombre de paramètres libres sous H_0 moins le nombre de paramètres libres sous H_1 , c'est-à-dire k .

L'hypothèse H_0 n'est pas rejetée à un niveau de signification α si $t < \chi^2_{k,\alpha}$, où $\chi^2_{k,\alpha}$ est la variable chi-2 à k degrés de liberté dont la probabilité au dépassement est α .

La statistique modifiée $t^* = (1-2.8/n)t$ approche plus efficacement la loi du chi-2 (Lawley, 1956). L'approximation est donc meilleure pour les échantillons de petite taille.

3.3.3 Le critère d'information d'Akaïke (CIA)

Le critère proposé par Akaïke (1974) permet de comparer des distributions n'appartenant pas nécessairement à la même famille. Il est défini de la manière suivante:

$$CIA = -2\ln L + 2p \quad (3.15)$$

où L est la fonction de vraisemblance de la distribution dont les p paramètres ont été estimés par la méthode du maximum de vraisemblance.

On remarque qu'en augmentant la complexité du modèle, le second terme de l'équation (3.15) augmente alors que le premier terme diminue à cause de l'amélioration de l'ajustement. La meilleure distribution donne le CIA minimum. L'annexe A traite plus en détail de l'origine du CIA.

Il existe une relation simple pour convertir les résultats du CIA en un test du rapport des vraisemblances maximales lorsqu'on compare deux distributions de même famille. Considérons la statistique t du test du rapport des vraisemblances maximales de l'équation (3.14) et les deux quantités suivantes déduites du CIA:

$$\ln L[H_0] = (2p - CIA[H_0])/2 \quad (3.16a)$$

$$\ln L[H_1] = (2(p+k) - CIA[H_1])/2 \quad (3.16b)$$

En substituant les quantités (3.16a) et (3.16b) dans l'équation (3.14), on obtient:

$$t = CIA[H_0] - CIA[H_1] + 2k \quad (3.17)$$

où $CIA[H_0]$ et $CIA[H_1]$ sont respectivement le CIA des lois F_0 et F_1 .

D'après la relation ci-dessus, on déduit facilement que le critère d'Akaïke est

équivalent à un test du rapport des vraisemblances maximales à un niveau de signification $P[\chi_k^2 \geq 2k] = \alpha$ dans le cas où F_0 et F_1 appartiennent à la même famille. En effet, d'après le critère d'Akaïke, la loi F_0 est préférée à la loi F_1 lorsque:

$$CIA[H_0] > CIA[H_1] \quad (3.18)$$

À partir des équations (3.17) et (3.18), on obtient alors l'inégalité suivante:

$$t > 2k \quad (3.19)$$

Ainsi, la plus petite valeur d' α qui ne rejete pas la loi F_0 est $P[\chi_k^2 \geq 2k] = \alpha$. Par exemple, lorsque $k=1$ on a $\alpha=16\%$ et lorsque $k=2$ on a $\alpha=14\%$.

3.3.4 Représentation graphique

La représentation sur un papier de probabilité approprié des valeurs de l'échantillon et de la distribution ajustée permet une évaluation visuelle de l'adéquation des données. Elle apporte également une comparaison subjective entre différentes lois ajustées par différentes méthodes lorsque les ajustements sont représentés sur le même graphique.

Pour les échantillons de petites tailles, les observations sont classées en ordre croissant $x_{(1)} \leq \dots \leq x_{(n)}$ et une probabilité empirique est attribuée à chacun des éléments. Les formules de probabilité empirique sont habituellement basées sur la moyenne, le mode ou la médiane de la fréquence observée. Cunnane (1978) a effectué une étude approfondie du sujet et suggère une formule compromis utilisable pour toutes les distributions:

$$P_k = (k-0.4)/(n-0.2) \quad (3.20)$$

Dans le cas des grands échantillons, la fréquence cumulée est attribuée à chacun des groupes formés.

Voici une liste des principaux types de papier de probabilité:

- Linéaire: La distribution uniforme apparaît comme une ligne droite avec une ordonnée linéaire.
- Normal: La distribution normale apparaît comme une ligne droite avec une ordonnée linéaire et la distribution log-normale comme une ligne droite avec une ordonnée logarithmique.
- Gumbel: La distribution Gumbel apparaît comme une ligne droite avec une ordonnée linéaire et la distribution Weibull comme une ligne droite avec une ordonnée logarithmique.
- Exponentiel: La distribution exponentielle apparaît comme une ligne droite avec une ordonnée linéaire.

CHAPITRE 4 REVUE DE LITTÉRATURE

Dans ce chapitre, nous identifierons premièrement les variables fondamentales de précipitation et leurs diverses applications pratiques. Par la suite, nous verrons les principales distributions statistiques utilisées dans la littérature pour représenter ces variables.

4.1 VARIABLES DE PRÉCIPITATION

Soit une échelle de temps discrète, par exemple journalière. Les phénomènes hydrométéorologiques de base intéressants à modéliser sont:

- L'occurrence des précipitations (jours secs et pluvieux);
- La durée des épisodes pluvieux (T_p) et des épisodes secs (T_s);
- La hauteur de pluie au cours d'une journée pluvieuse (H);

À cette liste s'ajoutent le maximum annuel de précipitation même si cette variable n'est pas une variable de base. La fréquence des extrêmes de pluie a des applications très importantes dans le dimensionnement des ouvrages hydrauliques et, pour cette raison, elle s'intègre bien à notre étude. La figure 4.1 est une représentation hypothétique de quelques unes des variables mentionnées précédemment.

En connaissant les lois qui régissent les variables de base, il est possible de déduire les lois gouvernant d'autres type de variables de précipitation. Citons entre autres la loi des hauteurs de pluie cumulées et la loi de la durée des cycles.

4.2 APPLICATIONS PRATIQUES

4.2.1 Variables de base

Les variables pluviométriques sont utilisées dans plusieurs domaines qui vont de l'agriculture à la construction des barrages en passant par le tourisme.

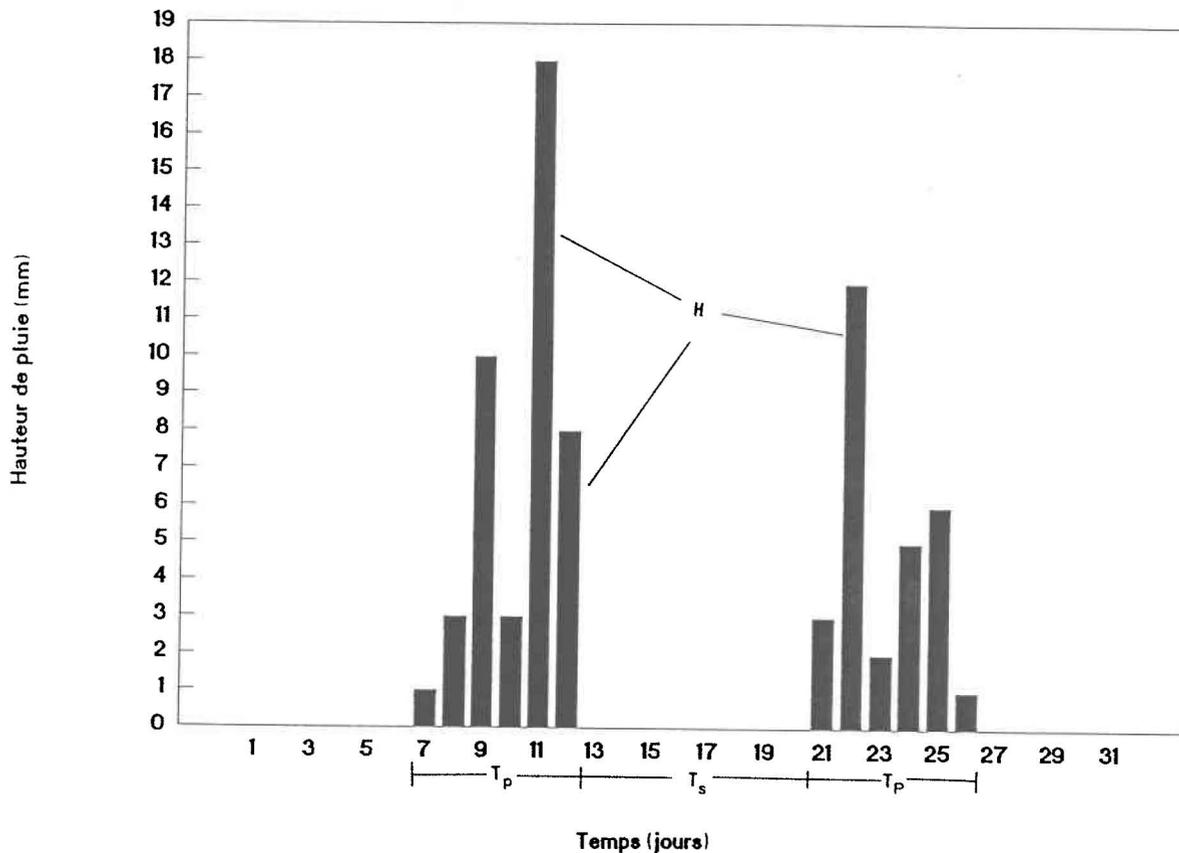


Figure 4.1 Hauteurs hypothétiques de pluie journalière au cours d'un mois

Dans les régions ayant un déficit en eau d'irrigation, l'agriculture nécessite la plus grande partie de l'eau disponible. La production des semailles est intimement liée à la quantité totale de pluie et aux épisodes pluvieux et secs durant la période de croissance. Une bonne compréhension de ces phénomènes permet de coordonner les politiques d'irrigation, d'étudier la capacité de production des sols et d'exploiter efficacement les fermes.

Au Canada, les troupeaux laitiers et les bovins de boucherie occupent une place remarquable dans l'industrie agricole. Le foin sert comme fourrage d'hiver à la majorité des animaux de ferme. Durant la période critique du séchage et de la récolte du foin, l'occurrence de certains phénomènes météorologiques dont une quantité minimale de pluie pendant un certain nombre de jours consécutifs est

nécessaire.

Les immenses forêts canadiennes ont un potentiel commercial immense. Par conséquence, il est donc primordial de protéger cette ressource. Un des buts premiers des données climatiques dans le domaine forestier est d'évaluer les risques d'incendies en forêts. Mentionnons, entre autres, l'importance de la quantité et de la durée des pluies ainsi que la période de temps séparant les précipitations utiles.

Les processus de bilan en eau tiennent compte des variables de base des précipitations. En particulier, la durée des épisodes secs (temps d'évapotranspiration) et des épisodes pluvieux (temps d'infiltration et de ruissellement) et la hauteur des précipitations (eau disponible à l'infiltration et au ruissellement) font parti du calcul du bilan en eau.

La limite des réservoirs et la capacité des aqueducs sont déterminées par la demande et la disponibilité en eau. Par exemple, pour prévenir une longue sécheresse, les réservoirs doivent être pleins pour faire face à la demande en eau des municipalités, de l'industrie et de l'agriculture. Les enregistrements mensuels et annuels de précipitation sont généralement examinés pour s'assurer d'une gestion adéquate des réservoirs.

Il existe plusieurs modèles informatiques de simulation des systèmes hydrologiques. Ces simulations rendent possible la quantification de plusieurs phénomènes reliés à l'hydrologie (par exemple, la prédiction de l'érosion et le transport des sédiments). Des séries synthétiques de pluie ayant les mêmes caractéristiques statistiques que les séries réelles de pluie servent souvent d'entrée à ces modèles.

De plus en plus, les systèmes de drainage urbain dans les grandes villes exigent une connaissance approfondie de la répartition temporelle de la pluie pour

des fins de conception, de gestion et d'opération. Il est alors important d'analyser les précipitations à des intervalles de temps de courte durée pour répondre à ces besoins.

Les précipitations ont également un rôle à jouer dans l'industrie du tourisme et des loisirs. Par exemple, les renseignements sur la fréquence des pluies guident les touristes dans la planification de leurs vacances et le choix de leurs vêtements et de leurs équipements.

4.2.2 Valeurs extrêmes de précipitation

Pour planifier la construction d'un barrage, la capacité de son réservoir, ses taux de débordement et les mesures qui en découlent (protection contre l'inondation), il est essentiel de connaître la fréquence, l'intensité et de la durée des précipitations extrêmes. Cet aspect du climat est important dans le design et la construction des toitures et de leurs gouttières, dans le choix de l'emplacement des constructions par rapport à la pente et à la stabilité du sol, dans la construction et l'entretien des routes, des ponts et des viaducs, dans l'orientation des rues par rapport à la topographie naturelle, etc...

Les maximums annuels de précipitation permettent de définir la notion de période de retour (ou temps de récurrence) T , souvent utilisée pour déterminer le dimensionnement des ouvrages hydrauliques (barrages, canaux d'irrigation, système de drainage urbain) et la lutte contre les inondations. Soit p la probabilité au dépassement de x_T :

$$P[X \geq x_T] = p = 1 - F(x_T) \quad (4.1)$$

alors, la période de retour de T année(s) est définie par la relation suivante:

$$1/T = p \quad (4.2)$$

La période de retour peut également être interprétée comme étant le nombre moyen d'années séparant des pluies telles que $x \geq x_T$.

Les ingénieurs ou les urbanistes choisissent la période de retour associée à une structure pour que le coût des dommages prévus lorsque la capacité de cette structure est dépassée (défaillance) soit inférieur au coût de son agrandissement. Naturellement, la protection du public est prise en compte. La période de retour dépend donc du genre d'ouvrage à construire, de son emplacement et des règlements gouvernementaux définis dans les codes de construction appliqués à l'échelle nationale, provinciale ou municipale.

Les données pluviométriques sont souvent utilisées pour connaître le débit associé à une période de retour donnée lorsqu'une rivière n'a pas de débitmètre. Il suffit de faire l'hypothèse qu'une pluie de fréquence donnée entraîne un débit de même fréquence (méthode rationnelle et autres procédures empiriques). Cependant, cette hypothèse n'est valable que pour un bassin versant homogène et une précipitation uniforme. À titre d'exemple, le tableau 4.1 indique les périodes de retour recommandées pour certaines constructions pour les rivières ayant un bassin versant inférieur à 20 km².

Dans les régions urbaines, la conception d'une structure pour le transport et le drainage des eaux de ruissellement est souvent conditionnée par la fréquence des pluies de courte durée. Au Canada, le réseau de drainage secondaire (par canalisation) est conçu en fonction de pluie de période de retour $T = 2$ à 10 ans, alors que le réseau principal (chaussées et canaux de surface) est conçu en fonction de pluie de période de retour $T = 25$ à 100 ans.

La durée de pluie utilisée correspond au temps de concentration (temps nécessaire pour atteindre l'équilibre) du bassin versant. Dans le cas d'un petit bassin la durée varie de 6 à 12 heures alors que dans une ville, elle joue de quelques minutes à quelques heures. Par exemple, dans le sud de l'Ontario, les réseaux d'égouts pluviaux des petites zones résidentielles sont construits pour faire face à des intensités de pluie d'une durée de 5 minutes et d'une période de retour

de 2 ans.

Tableau 4.1 Période de retour recommandée au Québec pour les bassins de 20 km² et moins (tiré de NRCC, 1989)

Situation	Période de retour recommandée (T)	
Dommages mineurs Plaine inondable restreinte Cultures extensives	5 ans	
Plaine inondable importante Cultures à hauts rendements Inondation de bâtiments de ferme	10 ans	
Routes:	Ponts	Ponceaux
rurales	25 ans	10 ans
principales	50 ans	25 ans
urbaines	100 ans	50 ans
autoroutes	100 ans	50 ans
chemins de fer	100 ans	100 ans

L'étude des extrêmes de précipitation a également d'autres applications dans des domaines tels l'érosion du sol, la perturbation des communications radio (courte durée) et le contrôle de la qualité des eaux.

4.3 OCCURENCE DES PRÉCIPITATIONS

L'occurrence des précipitations est étudiée selon deux approches différentes:

- a) **Modèle de Poisson:** on fait l'hypothèse que les événements secs et pluvieux sont indépendants. Cette hypothèse est applicable pour des orages de courtes durées (de l'ordre de l'heure).
- b) **Chaîne de Markov:** on fait l'hypothèse que les événements secs et pluvieux sont dépendants. Par exemple, cette hypothèse se destine aux dépressions dont le cycle de vie est de longue durée (plusieurs jours).

Lorsque les événements secs et pluvieux sont considérés indépendants et identiquement distribués sur une échelle de temps discret, on montre facilement que le nombre d'événements pluvieux sur une période donnée suit une distribution

binomiale (Smith et Schreider, 1973; Waymire et Gupta, 1981). Quoique la distribution binomiale ne soit pas très utilisée, son analogue sur l'échelle de temps continue, le processus de Poisson, est employée très fréquemment (Eagleson, 1978, 1981; Ozturk, 1981; Bardsley, 1984; Bonser et al., 1985).

Le principal désavantage de la distribution binomiale et du modèle de Poisson est dû à l'hypothèse d'indépendance des événements secs et pluvieux. Gabriel et Neumann (1962) ont montré que pour Tel-Aviv cette hypothèse ne pouvait pas être retenue. Ces auteurs ont pris en compte la dépendance des événements secs et pluvieux en considérant une chaîne de Markov du premier ordre à deux états. Plusieurs auteurs ont utilisé ce modèle pour expliquer l'occurrence des précipitations dans diverses régions du monde (Smith et Shreider, 1973; Buishand, 1978; Masson, 1980; Nguyen et Rousselle, 1981; Roldan et Woolhiser, 1982; Nguyen 1984a, 1984b). Chin (1977) a eu recourt à des chaînes de Markov d'ordre n , ce qui suppose que l'état actuel dépend des n états précédents, pour modéliser l'occurrence des précipitations aux États-Unis.

4.4 DURÉE DES ÉPISODES PLUVIEUX ET SECS

Il y a deux façons de définir les épisodes secs et pluvieux. Ils sont construits sur les bases:

- D'une série continue d'événements pluvieux bornée de deux événements secs pour l'épisode pluvieux (définition similaire pour l'épisode sec);
- D'un temps critique minimal qui sépare les épisodes pluvieux afin de regrouper ensemble tous les événements pluvieux ayant la même origine.

En générale, les épisodes pluvieux et secs ainsi construits sont non-corrélés et indépendants (Waymire et Gupta, 1981). Quélenec (1973) de même que Rao et Chenchayya (1974) rapportent des coefficients de corrélation d'ordre un faibles (moins de 0.1) pour la durée des épisodes secs et celle des épisodes pluvieux.

Les histogrammes des durées pluvieuses et sèches observées suggèrent un ajustement par des lois dont la densité décroît à mesure que la durée augmente. Le tableau 4.2 indique les distributions discrètes et continues les plus utilisées pour ajuster les durées des épisodes pluvieux et secs.

Tableau 4.2 Distributions utilisées pour modéliser les durées des épisodes pluvieux et secs

Durée	Auteurs	Région	Distributions
10 min.	Raudkivi et Lawgun, 1970 Rao et Chenchayya, 1974	Nouvelle-Zélande Indiana, E.-U.	WE WE, EX, GA
horaire	Grayman et Eagleson, 1969 Rao et Chenchayya, 1974 Egbuniwe, 1975 Bonser et al., 1985	Boston, E.-U. Indiana, E.-U. Nigéria Ontario, Canada	EX (T_p) EX, WE, GA EX (T_p), WE (T_s) EX, WE, GA
journalier	Gabriel et Neumann, 1962 Lobert, 1967 Hershfield, 1970 Quélenec, 1970 Quélenec, 1973 Buishand, 1978 Eagleson, 1978 Masson, 1980 Alexander, 1981 Roldan et Woolhiser, 1982 Revfiem, 1984	Israël France Denver, E.-U. France France Hollande Californie, E.-U. France Washington, E.-U. E.-U. Nouvelle-Zélande	GEO GEO GEO GEO EX, GA GEO, BNT, SL EX GEO (T_p) EX, WE, GA (T_s) EX BNT, GEO EX (T_p)

Voici une brève description de quelques unes de ces distributions:

GÉOMÉTRIQUE: Si la succession des événements secs et pluvieux est décrite par une chaîne de Markov du 1^{er} ordre à deux états on peut montrer que les durées des épisodes pluvieux et secs sont indépendantes et suivent des distributions géométriques (Waymire et Gupta, 1981 et section 5.6).

EXPONENTIELLE: Sur une échelle de temps continue la distribution géométrique tend vers la distribution exponentielle (Rice, 1987). Cette distribution est souvent employée en première approximation car elle est simple (1 paramètre). La propriété "sans-mémoire" de la loi exponentielle est irréaliste d'après Rao et Chenchayya (1974) puisque la probabilité de la fin d'un orage dépend de la durée passée de l'orage. Ces auteurs proposent l'utilisation des lois Weibull et Gamma pour éviter cet inconvénient.

WEIBULL et GAMMA: Ces distributions plus flexibles (2 paramètres) ajustent mieux certaines durées. Il est à noter que lorsque la durée n'est pas distribuée selon la loi exponentielle la succession des événements pluvieux et secs n'est plus un processus markovien (Rao et Chenchayya, 1974).

Rao et Chenchayya (1974) ont préféré la distribution Weibull aux distributions exponentielle et gamma pour représenter les durées des épisodes pluvieux et secs. Masson (1980) remarque qu'une loi géométrique suffit pour les épisodes pluvieux mais que les épisodes secs nécessitent une loi à deux paramètres comme la loi Weibull ou la loi gamma.

Il y a une assez forte corrélation linéaire entre la durée et la quantité de pluie tombée au cours d'un orage. Par exemple, Quénellec (1973) identifie des coefficients de corrélation de près de 0.8 pour deux bassins en France. Les résidus transformés de la régression linéaire sont ajustés par une loi bêta ou Weibull (Egbuniwe, 1975; Rao et Chenchayya, 1974). Une représentation alternative du lien entre la durée d'un épisode pluvieux et la hauteur de pluie est une distribution bivariée comme la distribution gamma bivariée.

4.5 HAUTEUR DE PRÉCIPITATION

Il existe deux approches dans la littérature pour modéliser statistiquement la hauteur de pluie tombée au cours d'un événement pluvieux:

- Les hauteurs successives de pluie non-nulle sont considérées indépendantes et une distribution est ajustée à l'échantillon.
- Les hauteurs successives de pluie non-nulle sont considérées indépendantes mais le lien entre les événements pluvieux et secs est ajouté au modèle. C'est-à-dire que la fonction de distribution dépend de l'événement précédent (sec ou pluvieux).

Les hauteurs de précipitation observées suggèrent que la dépendance des

hauteurs de précipitation successives augmente lorsque l'intervalle de temps diminue. Ainsi, Masson (1980), Quélenec (1973) et Buishand (1978) rapportent des coefficients d'autocorrélation d'ordre un variant de 0.1 à 0.2 pour une durée de 24 heures. En général, les pluies journalières successives non-nulles sont considérées indépendantes. En ce qui concerne les courtes durées, Rao et Chanchayya (1974) ont trouvé une autocorrélation d'ordre un significative entre les hauteurs de précipitation successives pour une période de 10 minutes alors que les données horaires ont été considérées indépendantes. Nguyen (1984b) a développé un modèle de répartition temporel d'un orage qui tient compte de la dépendance des événements secs et pluvieux et également de l'autocorrélation entre les hauteurs de pluie successives.

Les histogrammes des hauteurs de précipitation lors d'un événement pluvieux montre une fréquence qui diminue avec la hauteur. De plus, les données de précipitation indiquent une préférence pour des distributions ayant une "lourde queue" à droite. Le tableau 4.3 montre les distributions les plus employées dans la littérature pour ajuster les hauteurs de précipitation. Voici une brève description de quelques unes d'entre elles:

EXPONENTIELLE: Cette loi est très employée pour décrire les hauteurs de précipitation. Sa popularité vient de sa simplicité et de sa capacité à bien ajuster les données. Pour ces raisons, elle est régulièrement utilisée pour représenter la hauteur de précipitation dans des modèles de précipitation plus complexes.

GAMMA: La distribution gamma a été intensivement utilisée dans plusieurs modèles théoriques pour les données de précipitation des événements pluvieux. Suzuki (1964) propose l'utilisation de la loi gamma généralisée à trois paramètres pour ajuster les données de précipitation avec des durées de temps variables.

KAPPA: Mielke (1973) propose la distribution Kappa à 2 et 3 paramètres pour expliquer la propriété de "lourde queue" de la distribution des hauteurs de précipitation.

EXPONENTIELLE MIXTE: Cette distribution à 3 paramètres est composée par la

somme de 2 distributions exponentielles. Elle est employée pour tenir compte de l'évidence physique que les précipitations proviennent de deux masses d'air différentes: les orages de convection et toutes les autres formes de pluie.

Tableau 4.3 Distributions utilisées pour modéliser les hauteurs de précipitation

Durée	Auteurs	Région	Distributions
Annuelle	Markovic, 1965	E.-U. et Canada	N, LN, LN3, GA, GG
	Linsley et al., 1975	Région aride	LN
	Rao et Chenchayya, 1974	Région humide	N
	Phien et Jivajivajah, 1984	Indiana, E.-U.	N
Journalière	Rao et Rao, 1986	Portugal	LP3, GT
	Lobert, 1967	Indiana, E.-U.	N, GA
	Quélenec, 1970	France	LF
	Quélenec, 1973	France	EX, GA
	Smith et Schreider, 1974	France	EX, WE, GA
	Buishand, 1978	Arizona, E.-U.	EXM
	Richardson, 1982	Hollande	GA
	Woolhiser et Roldan, 1982	E.-U.	EX, GA, EXM
	Nguyen, 1984a	E.-U.	EX, GA, EXM
	Nguyen et Mayabi, 1990	Montréal, Canada	EX
Horaire	Nguyen et Rousselle, 1981	Montréal, Canada	EX, GA, WE, EXM
	Nguyen, 1984b	Montréal, Canada	EX
	Nguyen, 1987	Montréal, Canada	EX, GA, EXM
Diverse	Suzuki, 1964	Japon	GG
	Bernier et Fandoux, 1970	France	LF
	Mielke, 1973	E.-U.	KA, KA3, GA, LN

Certains auteurs utilisent une série de Fourier pour tenir compte de la variation saisonnière des paramètres (Woolhiser et Pegram, 1979; Woolhiser et Roldan, 1982).

Dans son étude sur les hauteurs de pluie journalière, Richardson (1982) a démontré la supériorité de la distribution exponentielle mixte à l'aide du test du rapport des vraisemblances maximales et de la représentation graphique. Woolhiser et Roldan (1982) arrivent à la même conclusion en utilisant le CIA et le test du rapport des vraisemblances maximales. Nguyen (1987) et Nguyen et Mayabi (1990) sont parvenus aux mêmes résultats pour les pluies journalières et horaires en ayant recours au test du chi-2, au CIA et à la représentation graphique. Il y a donc une tendance réelle en faveur de la distribution exponentielle mixte pour représenter les

hauteurs de précipitation journalière.

Stidd (1952) a suggéré que les hauteurs de précipitation élevées à une puissance fractionnaire ($1/2$ ou $1/3$) étaient distribuées normalement. Lavabre (1980) a effectué une transformation racine carrée des quantités de précipitation mensuelle et annuelle pour retenir la loi normale. Linsley et al. (1975) rapporte également une transformation racine carrée ou cubique pour normaliser les hauteurs de précipitation.

Pour les intervalles de temps de longue durée (mensuelle, saisonnière et annuelle), la loi normale est généralement retenue pour décrire la hauteur de précipitation. En effet, les totaux de précipitation sont asymptotiquement distribués selon la loi normale (sommes de variables aléatoires). Bardsley (1984) a démontré que les totaux de précipitation dans le cas d'un échantillon de taille fixée ont une asymétrie positive et un coefficient d'aplatissement supérieur à trois quelle que soit la distribution d'origine. La distribution des totaux est souvent déduite d'une loi régissant la quantité de précipitation (exponentielle, gamma) et d'un processus de Poisson (Quélenec, 1973; Eagleson, 1981, 1978; Ozturk, 1981).

4.6 VALEURS EXTRÊMES DE PRÉCIPITATION

La plupart des recherches dans le domaine des valeurs extrêmes en hydrologie utilisent trois différentes approches:

- L'analyse des séries annuelles de maxima
- L'analyse des séries de durée partielle
- L'analyse des séries chronologiques

Ici, nous nous attarderons plus spécialement aux séries annuelles de maxima et aux séries de durée partielle. La première est utilisée pour des événements à caractère exceptionnel alors que la seconde concerne des événements à caractère répétitif.

4.6.1 Séries annuelles de maxima

Comme son nom l'indique, dans les séries annuelles de maxima on considère la valeur maximale de pluie pour chaque année d'enregistrement. Il arrive plus rarement que la série soit définie sur une saison ou un mois plutôt que sur l'année. Les durées considérées varient de 5 minutes à plusieurs jours mais la durée de 24 heures est de loin la plus étudiée. En général, pour des raisons pratiques, la meilleure loi est déterminée pour les extrêmes journaliers et la même loi est utilisée pour toutes les autres durées.

Les données peuvent être de deux types: le maximum réel ou le maximum sur un intervalle d'observation à heure fixe. Ce dernier cas a le fâcheux inconvénient de couper en deux une pluie qui ne se produit pas entièrement dans un intervalle de temps d'observation fixe. Le U.S. Weather Bureau (1961) a développé une relation empirique pour ajuster les valeurs obtenues par une analyse statistique des maximums sur un intervalle d'observation fixe à celles obtenues par une analyse des maximums réels. Par exemple, le facteur de conversion pour une période de 24 heures est 1.13.

Comme on peut le remarquer au tableau 4.4, la distribution Gumbel est largement la plus utilisée pour représenter les maximums annuels. Les lois log-normale, Pearson type 3, log-Pearson type 3, log-Gumbel et GEV sont moins employées pour l'estimation des extrêmes de précipitation. Voici une courte description de quelques unes de ces lois:

GUMBEL: Le fait que la loi Gumbel est déduite comme loi limite de la valeur maximale de n variables aléatoires indépendantes et identiquement distribuées donne une justification théorique à son utilisation. Cependant, les hypothèses conduisant à la distribution de Gumbel ne sont pas toujours respectées:

- les hauteurs de pluie ne sont pas nécessairement indépendantes entre elles (en particulier pour les courtes durées);
- la taille de chaque échantillon n'est pas nécessairement très grande (par

exemple, $n=365$ pour les hauteurs de pluie journalière);

- les hauteurs de pluie subissent des variations saisonnières.

En pratique, la loi Gumbel peut quand même être appliquée comme toute autre distribution.

LOG-NORMALE: Chow (1962) a offert une interprétation théorique pour l'emploi de la distribution log-normale aussi connue sous le nom de loi de Gibrat-Gauss (Gibrat, 1931). Il explique que l'occurrence d'un événement hydrologique est le résultat du produit de plusieurs facteurs indépendants. Par le théorème de la limite centrale, on peut montrer que le logarithme du résultat est distribué normalement.

LOG-PEARSON TYPE 3: Cette distribution a été recommandée par le Conseil des Ressources en Eau des États-Unis (Benson, 1968) pour l'estimation des crues.

Tableau 4.4 Distributions utilisées pour modéliser les séries annuelles de maxima

Auteurs	Région	Distributions
U.S Weather Bureau, 1961	E.-U.	GU
Huff et Neil, 1959	Illinois, E.-U.	GU, LGU, LN
Reich, 1963	Afrique du Sud	GU
Majumbar et Sawhney, 1965	Inde	GU, LN, LP3
Stol, 1971	Hollande	GU
Samuelson, 1972	Crète, Grèce	GU, LN
Miller et al., 1973	E.-U.	GU, P3, LP3
Cao, 1974	Sardaigne, Italie	LN
Dickinson, 1977	Ontario, Canada	GU
Baghiratham et Shaw, 1978	Sri Lanka	GU
CTGREF, 1978	France	GU, LGU, LN, LP3
Kappus et al., 1978	Iran	GU
Dubant et al., 1980	France	GU, P3, LN
Masson, 1980	France	GU
Lahaye, 1980	Haute-Volta	GU
Revfiem, 1983	Nouvelle-Zélande	GU
Hogg et Carr, 1985	Canada	GU
Sharma, 1987	Zambie	GU, LN, P3, LP3, GEV
Buishand et Demaré, 1990	Belgique	GU

Buishand et Demaré (1990), Dickinson (1977), Stol (1971) se sont attardé aux extrêmes mensuels de pluie pour prédire les extrêmes annuels de pluie. Cette approche est située à mi-chemin entre les séries annuelles de maxima et les séries de durée partielle.

La corrélation entre les maximums annuels de pluie est généralement faible et la plupart des auteurs ne vérifient pas cette hypothèse.

4.6.2 Séries de durée partielle

Les séries de durée partielle sont composées de l'ensemble des valeurs de pluie supérieure à un seuil fixé. La méthode pour analyser ces séries consiste à considérer la distribution du nombre de dépassements (valeur supérieure au seuil) et la distribution de la grandeur du dépassement. Le tableau 3.4 montre que le modèle de poisson et la loi exponentielle sont les plus fréquemment employés.

Tableau 4.5 Distributions utilisées pour modéliser les séries de durée partielle

Auteurs	Région	Distributions	
		# dépassement	Dépassement
Thom, 1959	E-U	Poisson	Exponentielle
Colin et Bedel, 1980	France	Poisson	Exponentielle
Revfiem, 1984	Nouvelle-Zélande	Poisson	Exponentielle
Van Montfort et Witter, 1986	Hollande	Poisson	Pareto généralisée
Zucchini et Adamson, 1989	Afrique du Sud	Poisson	Exponentielle
Fitzgerald, 1989	Irlande	Binomial négative	Gamma
Buishand et Demaré, 1990	Belgique	Poisson	Pareto généralisée
		Poisson	Exponentielle

L'avantage d'utiliser les séries de durée partielle par rapport aux séries annuelles de maxima est d'augmenter le nombre d'éléments considérés dans l'échantillon. Malheureusement, il y a deux principaux problèmes concernant cette approche. Le premier est l'indépendance des dépassements: par exemple, plusieurs valeurs excédant la valeur de base proviennent de la même condition météorologique. En second lieu, il n'existe pas de méthode systématique pour déterminer le seuil de référence (différents seuils conduisent à différents résultats).

Le U.S. Weather Bureau (1961) a développé une relation empirique entre les valeurs extrêmes calculées à partir des séries annuelles de maxima et celles des séries de durée partielle. Cette étude montre que les calculs basés sur les maxima annuels donnent des résultats semblables à ceux basés sur les séries de durée partielle pour des périodes de retour supérieures à 10 ans. Pour les périodes de retour plus courtes, les résultats des séries de durée partielle sont supérieurs à ceux obtenus à partir des séries annuelles de maxima. Les facteurs de conversion

appropriés sont présentés au tableau 4.6.

Tableau 4.6 Facteur de conversion entre les résultats basés sur les séries annuelles de maxima et ceux basés sur les séries de durée partielle

Période de retour	Facteur de conversion
2 ans	1.13
5 ans	1.04
10 ans	1.01

4.7 AUTRES PHÉNOMÈNES MÉTÉOROLOGIQUES

Plus rarement que pour les précipitations, les météorologues trouvent parfois nécessaire d'étudier la probabilité d'autres événements climatiques. Linsley et al. (1982) et Suzuki (1980) ont résumé l'état de la recherche dans ce domaine. Cette section, discute brièvement de l'ajustement de ces autres variables météorologiques.

VITESSE DU VENT: Que ce soit pour la vitesse instantanée, la moyenne horaire ou journalière, il n'existe pas de distribution satisfaisante pour ajuster ces variables. En effet, les histogrammes montrent une distribution bimodale avec un maximum près de zéro et un autre à une certaine valeur plus élevée. En ce qui concerne le maximum annuel, la distribution Gumbel est habituellement utilisée.

DIRECTION DU VENT: Les modèles de distribution empirique circulaire ou normale circulaire sont généralement employés pour représenter la rose des vents.

TEMPÉRATURE: Le maximum estival annuel et le minimum hivernal annuel se conforment raisonnablement bien à la distribution normale. Parfois, le maximum estival annuel a une asymétrie positive tandis que le minimum hivernal a une asymétrie négative. La distribution Pearson type I (Bêta) est aussi utilisée.

PRESSION ATMOSPHÉRIQUE: La distribution de la pression atmosphérique par saison est quasi-normale avec quelquefois une légère asymétrie négative.

HUMIDITÉ RELATIVE: La distribution de l'humidité relative a une asymétrie positive. La somme de deux normales et la Pearson type 1 conduisent à un ajustement raisonnable.

CHAPITRE 5 ÉPISODES PLUVIEUX ET SECS

5.1 CHOIX DES STATIONS

La plupart des stations du réseau pluviométrique du Service de l'environnement atmosphérique (SEA) permettent de mesurer les hauteurs de précipitation quotidienne. Les données climatiques sont disponibles sous une forme archivée et dans un format qui se prête bien à un traitement informatique. Les stations et le type de variables météorologiques disponibles sont répertoriés dans les catalogues des stations climatologiques établis pour les différentes régions canadiennes.

La sélection des stations pour l'étude des épisodes secs et pluvieux est basée sur les critères suivants:

- une diversité des régions géographiques et climatiques pour bien représenter les provinces du Québec et du Nouveau-Brunswick;
- une période d'enregistrement récente avec un minimum de données manquantes.

La liste et les principales caractéristiques des stations choisies se trouvent au tableau 5.1. Dans le but d'alléger le texte et les tableaux, les stations sont désignées par une abréviation composée de deux lettres (la province) et d'un chiffre (première colonne du tableau 5.1).

La longueur des enregistrements pour les 8 stations varie entre 24 et 36 ans. La période d'enregistrement durant l'année s'étale du mois de juin au mois de septembre (saison estivale). Ce choix est motivé par le désir de ne pas inclure le phénomène des précipitations solides (neige) avec celui des précipitations liquides (pluie) et d'éliminer, si possible, l'effet de la variation saisonnière.

Les hauteurs de pluie sont mesurées à l'aide d'un pluviomètre standard en cuivre. Une éprouvette en verre graduée sert à déterminer la hauteur de l'eau

recueillie dans le récipient interne du pluviomètre. La précision de la lecture est de l'ordre de 0.2 mm. Une pluie inférieure à 0.2 mm n'est pas mesurée la journée même, mais elle est ajoutée à la journée suivante. De si petites quantités d'eau peuvent être causé par le brouillard ou la rosée. D'un autre coté, de petites valeurs de pluie peuvent être enregistrées comme étant nulle. On voit bien que la limite de détection du pluviomètre est à l'origine d'incertitude dans le processus d'occurrence de la pluie.

Tableau 5.1 Stations sélectionnées pour l'étude des durées des épisodes pluvieux et secs

Station	Code	Ville	Période d'observation
QU1	7028124	Sherbrooke	1962-1990
QU2	7047910	Sept-Iles	1960-1990
QU3	7090120	Amos	1963-1990
QU4	7010720	Berthierville	1966-1990
QU5	7040440	Baie Comeau	1964-1990
NB1	8100880	Charlo	1967-1990
NB2	8101000	Chatam	1960-1990
NB3	8104900	St-John	1955-1990

La mesure de la hauteur de pluie est l'objet d'une certaine imprécision qui peut être due à la défaillance d'un instrument ou à une erreur d'observation. Le vent est, cependant, la principale source d'erreur systématique. Il est généralement responsable de valeurs de 5 à 15 % inférieures aux valeurs réelles (Sevruk et Geiger, 1981).

5.2 DÉFINITION DES ÉPISODES

Un jour pluvieux est défini comme étant un jour ayant une hauteur de pluie supérieure à un certain seuil (habituellement égal au seuil de détection de l'instrument). De son côté, un jour sec est défini comme étant un jour ayant une hauteur de pluie inférieure au seuil considéré.

Si on note P_j l'état pluvieux du jour j et S_j l'état sec du jour j , un épisode sec de longueur k est défini par une série continue de k jours secs bornée de deux jours pluvieux:

$$\{P_{i-1}S_iS_{i+1}, \dots, S_{i+k-1}P_{i+k}\}$$

L'épisode pluvieux de longueur m est définie de manière analogue par:

$$\{S_{i-1}P_iP_{i+1}, \dots, P_{i+m-1}S_{i+m}\}$$

Il est évident que différents seuils vont conduire à différents résultats. Le seuil dépend de l'utilisation future de nos résultats. Par exemple, un seuil bas est important pour rendre compte des besoins en eau de certaines espèces de plantes alors qu'un seuil plus élevé est approprié pour prédire la production en eau de grands bassins ou pour modéliser le ruissellement. Dans cette étude une journée est considérée pluvieuse si la hauteur de pluie est non-nulle (en réalité, supérieure ou égale à 0.2 mm).

Les échantillons des durées des épisodes pluvieux et secs utilisés dans ce chapitre sont donc construites sur la base des deux définitions ci-dessus à partir des enregistrements des hauteurs de pluies journalières.

5.3 CRITÈRES ET TESTS STATISTIQUES

Les échantillons des durées des épisodes pluvieux et secs doivent satisfaire plusieurs critères statistiques. Dans cette section, on s'attarde à l'examen des hypothèses d'indépendance, d'homogénéité et d'existence de valeurs singulières.

5.3.1 Indépendance

L'hypothèse d'indépendance de la durée des épisodes secs et pluvieux est vérifiée à partir du test de Wald-Wolfowitz (section 3.2.1). La statistique $|u|$ du test est calculée pour chaque année d'enregistrement plutôt que pour l'ensemble de l'échantillon puisque que les observations ne sont pas considérées durant l'hiver (il y a une coupure à chaque année). En raison du grand nombre d'années impliqués,

seulement celles dont la valeur de $|u|$ dépasse la valeur critique à un niveau de signification de 5% se trouvent au tableau 5.2a pour la durée des épisodes pluvieux et au tableau 5.2b pour la durée des épisodes secs. Rappelons que la valeur critique à un niveau de signification de 5% est 1.96 et à 1% de 2.57. À titre indicatif, le coefficient d'autocorrélation d'ordre un ρ_1 (section 3.1.1) pour les épisodes pluvieux et les épisodes secs est inclus à la dernière colonne des deux tableaux.

Tableau 5.2a La statistique $|u|$ du test d'indépendance de Wald-Wolfowitz pour les épisodes pluvieux (rejet à 5%)

Station	Année	$ u $	ρ_1
QU1	1966	2.07	-0.45
	1975	2.67*	0.66
	1979	2.43	0.41
QU3	1967	2.32	0.34
	QU4	1977	2.97*
		1985	2.08
QU5	1978	2.41	0.37
	1979	2.11	-0.40
NB1	1973	2.17	0.29
NB2	1973	2.13	0.34
	1977	1.99	0.25
NB3	1961	2.01	0.36
	1979	2.31	0.29

* rejet à 1%

Tableau 5.2b La statistique $|u|$ du test d'indépendance de Wald-Wolfowitz pour les épisodes secs (rejet à 5%)

Station	Année	$ u $	ρ_1
QU1	1962	2.25	0.30
QU2	1987	2.06	0.33
QU3	1979	2.35	0.43
QU4	1986	2.13	0.33
QU5	1965	2.73*	0.29
NB2	1988	2.38	0.37
NB3	1961	2.06	0.37

* rejet à 1%

La figure 5.1 représente deux graphiques "boîtes à moustaches" (Rice, 1987) du coefficient d'autocorrélation d'ordre un calculé à chaque année pour la durée des épisodes pluvieux et secs. Voici une brève description de ce genre de graphique:

- la boîte centrale contient 50% des valeurs situées entre le troisième quartile (75%) et le premier quartile (25%). La distance entre le premier et le troisième quartiles est la distance interquartile (DIQ);
- la ligne centrale représente la médiane;
- les "moustaches" s'étendent jusqu'à la valeur la plus éloignée à l'intérieur de la distance 1.5 DIQ du premier quartile et du troisième quartile;
- les points sont des valeurs "extrêmes" situées à l'extérieur des "moustaches" mais à l'intérieur d'une distance de 3.0 DIQ de l'un ou l'autre des quartiles et les plus (+) sont des valeurs "exceptionnelles" situées à une distance supérieures à 3.0 DIQ.

La figure 5.1 donne ainsi une idée de l'étalement de ρ_1 .

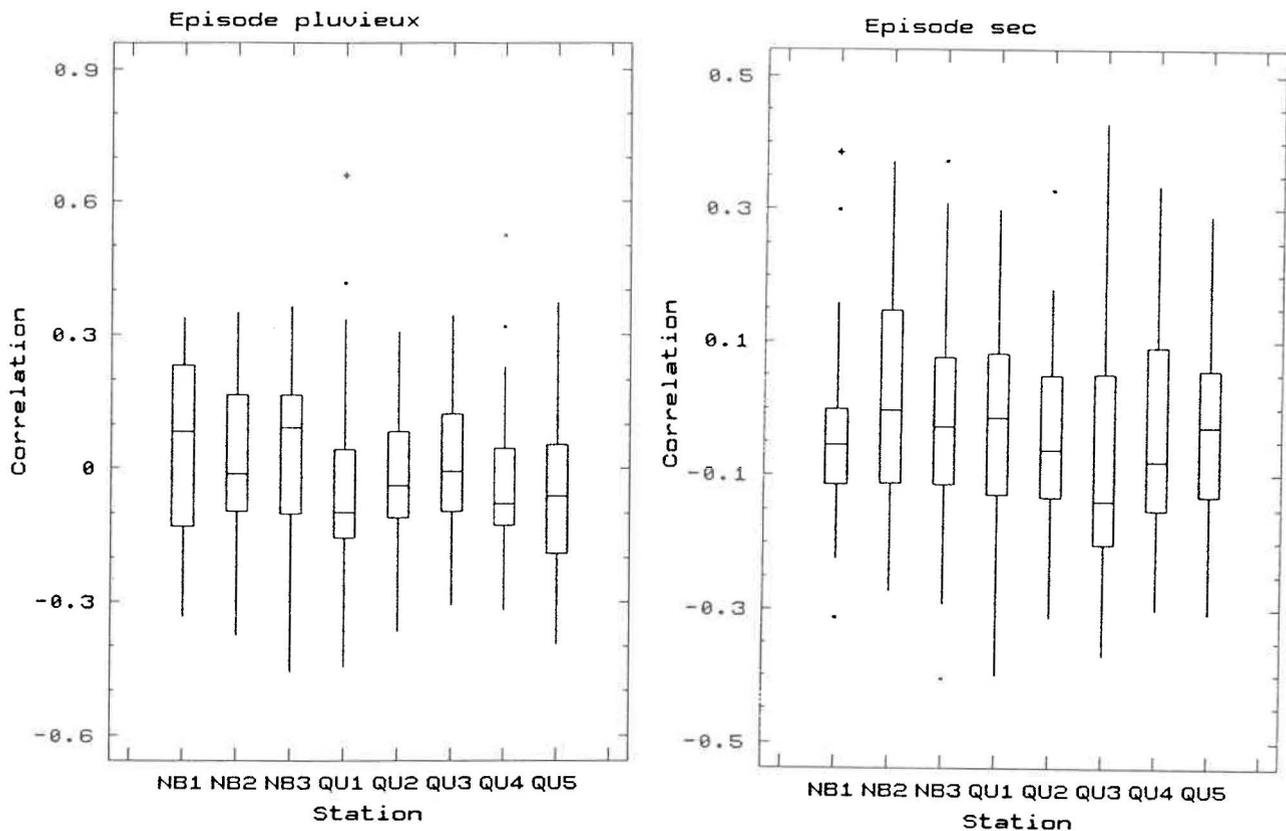


Figure 5.1 Boîtes à moustaches de l'autocorrélation d'ordre un des durées des épisodes pluvieux et secs

À la lumière des tableaux 5.2a et 5.2b, la présence d'années ayant une

dépendance remarquée se produit à l'occasion. D'un point de vue statistique, ce résultat n'est pas surprenant car à un niveau de signification de 5% (1 cas sur 20), on peut s'attendre à rejeter l'hypothèse d'indépendance alors qu'elle est vraie (erreur de type I) 12 années sur 231 (nombre d'années pour l'ensemble des stations) et 2 années sur 231 à un niveau de signification de 1% (1 cas sur 100). Ces chiffres correspondent approximativement à ce qui est observé aux tableaux 5.2a et 5.2b.

La figure 5.1 montre des médianes d'autocorrélation d'ordre 1 voisines de zéro. De plus, le coefficient de corrélation est parfois positif et parfois négatif ce qui est difficile à expliquer physiquement, sauf par une variabilité de ρ , autour de 0. Ainsi, l'hypothèse d'indépendance des épisodes secs successifs et des épisodes pluvieux successifs peut être acceptée pour l'ensemble des stations considérées.

Dans le but de vérifier la relation linéaire entre la longueur des épisodes pluvieux (T_p) et la hauteur de pluie tombée lors de cet épisode (H_p), le coefficient de corrélation entre H_p et T_p est calculé pour chacune des stations. Des valeurs variant entre 0.54 et 0.64 indiquent la possibilité d'un lien linéaire entre ces deux variables.

5.3.2 Homogénéité

L'homogénéité chronologique n'est pas vérifiée en raison de la bonne qualité des enregistrements (pas d'interruption d'exploitation de la station, pas de déplacement de la station). Également, la période d'enregistrement relativement courte (en moyenne 29 ans) ne nous permet pas d'identifier d'année de coupure ni de changements climatiques ou de cycles à long terme.

En ce qui à trait à l'homogénéité saisonnière, le test de Kruskal-Wallis (section 3.2.3) est appliqué aux différentes stations. En général, les précipitations subissent une variation saisonnière en raison du déplacement annuel des grands systèmes atmosphériques. Les échantillons de la durée des épisodes sont donc divisés en quatre sous-échantillons (juin, juillet, août, septembre) selon le mois durant lequel

se produit le début de l'épisode afin d'identifier ces hétérogénéités. Le tableau 5.3 présente les valeurs de la statistique τ .

En général, les durées des épisodes pluvieux et secs peuvent être considérées homogènes lors de la saison estivale allant de juin à septembre et peuvent donc être regroupées. Une exception s'observe pour la station NB3 (épisodes secs) qui a une valeur élevée de τ . Une étude plus approfondie de cet échantillon montre qu'une division en deux sous-échantillons {juin} et {juillet,août,septembre} est appropriée. Dans la suite de l'étude, l'indicateur du sous-échantillon {juin} est NB3A et celui du sous-échantillon {juillet,août,septembre} est NB3B.

Tableau 5.3 La statistique τ du test d'homogénéité mensuelle de Kruskal-Wallis pour les épisodes pluvieux et secs

Station	τ	
	Épisodes pluvieux	Épisodes secs
QU1	1.13	5.93
QU2	1.21	3.78
QU3	4.91	4.12
QU4	4.96	3.09
QU5	1.74	0.50
NB1	2.50	1.66
NB2	4.77	5.54
NB3	2.26	11.02*

* rejet à 5%

5.3.3 Valeurs singulières

Le test de Grubbs-Beck (section 3.2.4) est appliqué à chacun des échantillons des durées des épisodes pluvieux et secs. En raison de l'asymétrie positive des durées, l'hypothèse de la log-normalité des échantillons peut être acceptée pour l'application du test. Les valeurs maximums observées pour les épisodes pluvieux sont en moyenne de 10 jours alors que pour les épisodes secs elles sont en moyenne de 15.5 jours. Aucune valeur singulière n'est à signaler à un niveau de signification de 1%. Cependant, soulignons que ce test est peu puissant et très

tributaire de l'hypothèse de log-normalité.

5.4 CARACTÉRISTIQUES STATISTIQUES DES ÉPISODES PLUVIEUX ET SECS

Afin d'avoir une vue d'ensemble des échantillons considérées, quelques moments sont calculés. Il s'agit de la moyenne (\bar{x}) et les coefficients de variation (\hat{C}_v), d'asymétrie (\hat{C}_s) et d'aplatissement (\hat{C}_k) décrits à la section 2.1.2. Les résultats sont présentés au tableau 5.4a pour les épisodes pluvieux et au tableau 5.4b pour les épisodes secs.

Tableau 5.4a Caractéristiques statistiques des épisodes pluvieux

Station	\bar{x} (jours)	\hat{C}_v	\hat{C}_s	\hat{C}_k	n
QU1	2.14	0.67	1.77	7.21	732
QU2	2.06	0.66	1.89	7.42	790
QU3	2.23	0.73	2.05	8.93	720
QU4	2.04	0.66	1.98	8.15	643
QU5	1.97	0.65	1.83	7.40	654
NB1	2.03	0.69	2.08	9.19	613
NB2	2.03	0.69	1.86	6.92	755
NB3	1.92	0.62	1.82	7.05	889
MOY	2.05	0.67	1.91	7.78	

Tableau 5.4b Caractéristiques statistiques des épisodes secs

Station	\bar{x} (jours)	\hat{C}_v	\hat{C}_s	\hat{C}_k	n
QU1	2.49	0.76	1.82	6.73	737
QU2	2.55	0.79	2.25	9.81	797
QU3	2.30	0.76	1.96	8.22	720
QU4	2.52	0.77	1.80	6.59	646
QU5	2.72	0.83	2.85	16.83	657
NB1	2.58	0.76	1.99	8.46	619
NB2	2.81	0.78	2.10	9.13	759
NB3A	2.43	0.78	2.02	7.85	206
NB3B	2.98	0.82	2.15	9.22	699
MOY	2.60	0.78	2.10	9.20	

On note que la durée moyenne des épisodes pluvieux est d'environ 2 jours tandis que pour celles des épisodes secs elle est d'environ 2.5 jours. Donc, les épisodes secs sont plus longs que les épisodes pluvieux dans l'est canadien durant l'été.

L'asymétrie des épisodes pluvieux et secs est positive et voisine de 2. Les épisodes secs sont en général plus asymétriques que les épisodes pluvieux.

Le coefficient d'aplatissement tant qu'à lui est toujours supérieur à 6 et est, habituellement, sous la valeur de 10. L'aplatissement semble moindre pour les épisodes pluvieux que les épisodes secs.

Finalement, la taille des échantillons varie entre 600 et 900 épisodes sauf en ce qui concerne la station NB3A pour les épisodes secs (231 épisodes).

5.5 AJUSTEMENT DES ÉPISODES PLUVIEUX ET SECS

5.5.1 Distributions et méthodes d'ajustement considérées

Les histogrammes des épisodes pluvieux et secs montrent une diminution du nombre d'observations à mesure que la longueur de la durée augmente (figure 5.2). De plus, l'échelle de temps est divisée en pas de 1 jour. Les lois et les méthodes d'estimation des paramètres utilisées dans l'étude sont les suivantes:

Géométrique (GEO):	Maximum de vraisemblance (MV)
Série logarithmique (SL):	Maximum de vraisemblance (MV)
Binomiale négative tronquée (BNT):	Maximum de vraisemblance (MV) Méthode des moments (MM)

Ces trois distributions sont généralement utilisées lors de l'ajustement de la durée des épisodes pluvieux et secs par des distributions discrètes.

5.5.2 Calculs effectués

Des programmes informatiques utilisant le langage GAUSS ont été développés pour estimer les paramètres des 3 différentes distributions considérées. Les tableaux 5.5a et 5.5b donnent les paramètres estimés pour respectivement les épisodes pluvieux et les épisodes secs.

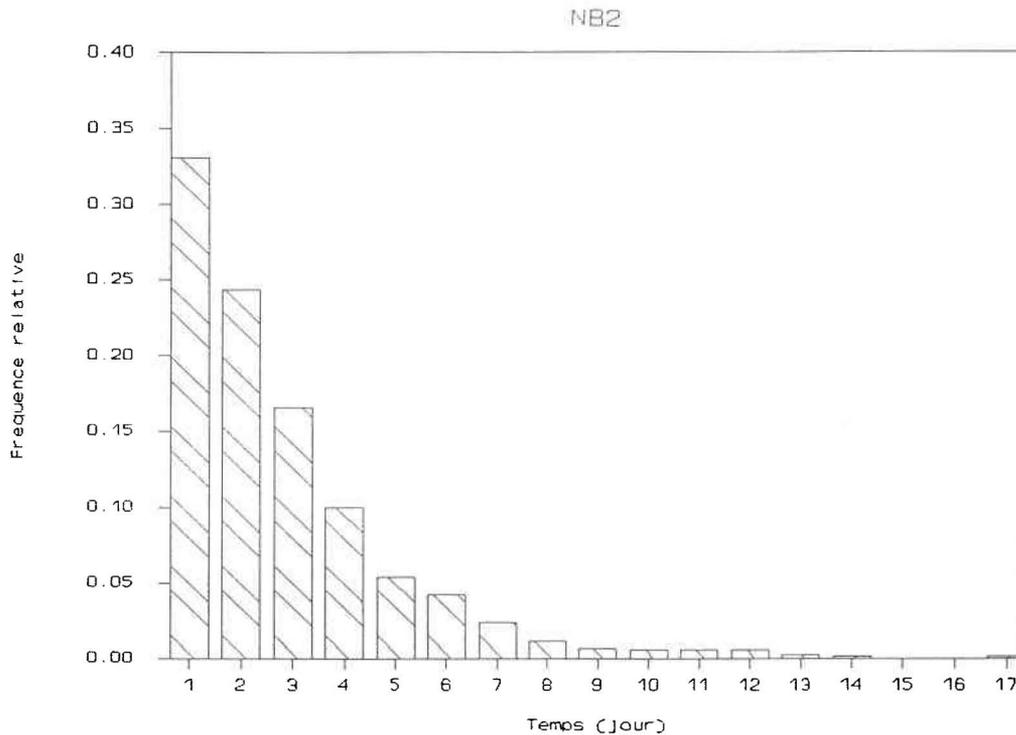


Figure 5.2 Histogramme typique de la durée des épisodes secs

Rappelons que la distribution géométrique et celle de la série logarithmique sont des cas particuliers de la distribution binomiale négative tronquée lorsque respectivement $r=1$ et $r \rightarrow 0$ (section 2.3.3). L'examen des tableaux 5.5a et 5.5b montre que la distribution série logarithmique est moins appropriée que la distribution géométrique puisque l'on a pour la loi binomiale négative tronquée une valeur de r près de 1. Également, la distribution géométrique est mieux adaptée à la durée des épisodes secs qu'à celle des épisodes pluvieux car r est plus rapproché de 1 dans le premier cas que dans le deuxième.

Tableau 5.5a Paramètres estimés pour les épisodes pluvieux (GEO, SL, BNT)

Station	GEO	SL	BNT			
	MV	MV	MV		MM	
	\hat{q}	$\hat{\theta}$	\hat{r}	\hat{b}	\hat{r}	\hat{b}
QU1	0.47	0.75	2.28	0.61	2.19	0.60
QU2	0.49	0.73	2.67	0.66	2.33	0.63
QU3	0.45	0.76	1.33	0.49	1.24	0.48
QU4	0.49	0.72	2.54	0.65	2.17	0.62
QU5	0.51	0.71	2.10	0.63	2.02	0.63
NB1	0.49	0.72	1.71	0.58	1.52	0.56
NB2	0.49	0.72	1.56	0.56	1.48	0.56
NB3	0.52	0.69	3.48	0.73	3.10	0.71
MOY	0.49	0.72	2.21	0.61	2.01	0.60

Tableau 5.5b Paramètres estimés pour les épisodes secs (GEO, SL, BNT)

Station	GEO	SL	BNT			
	MV	MV	MV		MM	
	\hat{q}	$\hat{\theta}$	\hat{r}	\hat{b}	\hat{r}	\hat{b}
QU1	0.40	0.80	1.07	0.41	1.09	0.41
QU2	0.39	0.81	1.26	0.43	1.06	0.40
QU3	0.43	0.77	0.95	0.43	0.97	0.43
QU4	0.40	0.80	0.99	0.40	1.04	0.40
QU5	0.37	0.83	1.29	0.41	0.96	0.35
NB1	0.39	0.81	1.38	0.44	1.29	0.43
NB2	0.36	0.83	1.39	0.41	1.24	0.39
NB3A	0.41	0.79	1.08	0.42	0.99	0.41
NB3B	0.34	0.85	1.21	0.36	1.07	0.34
MOY	0.39	0.81	1.18	0.41	1.08	0.40

5.6 LIEN ENTRE LA CHAÎNE DE MARKOV ET LA LOI GÉOMÉTRIQUE

Nous avons vu dans la revue de littérature (section 4.3) que l'état (sec ou pluvieux) des jours successifs sont considérés dépendants (chaîne de Markov) ou indépendants (distribution binomiale ou modèle de Poisson). Naturellement, la durée des épisodes pluvieux et secs est directement liée à cette supposition. Cette section

a pour but de déduire la distribution de la durée des épisodes pluvieux et secs dans l'hypothèse d'un processus de Markov homogène d'ordre un à deux états.

Si l'état du jour $i+1$ ne dépend que de l'état du jour i , le processus d'occurrence de la pluie est complètement déterminé par la matrice de transition (les probabilités conditionnelles de l'état du jour $i+1$ sachant l'état du jour i):

Jour i	Jour $i+1$		Somme
	Sec	Pluvieux	
Sec	$1-q_s$	q_s	1
Pluvieux	q_p	$1-q_p$	1

où q_p est la probabilité que le jour $i+1$ soit sec sachant que le jour i est pluvieux;
 q_s est la probabilité que le jour $i+1$ soit pluvieux sachant que le jour i est sec.

Puisque le processus est homogène, les probabilités q_s et q_p sont invariants quelque soit le jour i . À partir de la définition des épisodes et de la matrice de transition, on déduit la loi de probabilité de la durée d'un épisode pluvieux:

$$P[T_p=k] = P[P_1 P_2 \dots P_k S_{k+1} | P_1] \quad (5.1a)$$

on a k jours pluvieux consécutifs et le jour $(k+1)$ est sec.

$$P[T_p=k] = P[(S_{k+1} | P_1) \text{ et } (P_k P_{k-1} \dots P_2 P_1 | P_1)] \quad (5.1b)$$

En utilisant la loi de multiplication (Rice, 1987), on obtient

$$P[T_p=k] = P[S_{k+1} | (P_k P_{k-1} \dots P_2 P_1)] \cdot P[P_k \dots P_1 | P_1] \quad (5.1c)$$

Puisque le processus est homogène et d'ordre 1:

$$P[S_{k+1} | (P_k \dots P_1)] = P[S_{k+1} | P_k] = P[S | P] \quad (5.1d)$$

On peut recommencer de la même manière avec:

$$P[(P_k \dots P_1) | P_1] = P[P_k | P_{k-1}] \cdot P[(P_{k-1} \dots P_1) | P_1] \quad (5.1e)$$

et en continuant, on obtient:

$$P[T_p=k] = P[S | P] \cdot (P[P | P])^{k-1} = q_p (1-q_p)^{k-1} \quad (5.1f)$$

On retrouve la loi géométrique décrite à la section 2.3.1. En effet, pour un épisode

pluvieux de durée k on observe $(k-1)$ couple (PP) consécutifs et un couple (PS).

D'une manière identique on peut montrer que la durée d'un épisode sec suit une distribution géométrique:

$$f(T_s=m) = q_s(1-q_s)^{m-1} \quad (5.1b)$$

On peut également déterminer les probabilités marginales, on a:

$$P[P_{i+1}] = P[(P_{i+1} \text{ et } P_i) \text{ ou } (P_{i+1} \text{ et } S_i)] \quad (5.2a)$$

En utilisant la loi de multiplication et d'addition (Rice, 1987), on obtient:

$$P[P_{i+1}] = P[(P_{i+1} | P_i) \cdot P[P_i] + P[P_{i+1} | S_i] \cdot P[S_i]] \quad (5.2b)$$

S'il y a homogénéité, on a quelque soit le jour i

$$P[P_{i+1}] = P[P_i] = P[P] \quad (5.2c)$$

La relation (5.2b) devient:

$$P[P] = (1-q_p)P[P] + q_s(1-P[P]) \quad (5.2d)$$

On en déduit:

$$P[P] = q_s / (q_s + q_p) \quad (5.2e)$$

$$P[S] = q_p / (q_s + q_p) \quad (5.2f)$$

où $P[P]$ est la probabilité d'avoir une journée pluvieuse;

$P[S]$ est la probabilité d'avoir une journée sèche. Évidemment, $P[P] + P[S] = 1$.

Dans le cas où l'on suppose que les événements secs et pluvieux sont considérés indépendants et identiquement distribués (binomiale), on trouve à partir de la matrice de transition la relation:

$$P[S|P] = P[S|S] \quad (5.3)$$

$$q_p = 1 - q_s$$

Ainsi, la loi de probabilité de la durée des épisodes est aussi une distribution géométrique dont les paramètres sont donnés aux équations (5.2e) et (5.2f). Donc, pour que l'hypothèse d'indépendance soit respectée, la somme de la première colonne du tableau 5.5a avec celle du tableau 5.5b doit donner 1. En réalité, cette

sommation est toujours inférieure à 0.9. Il est donc possible de conclure qu'il y a une dépendance entre les événements secs et pluvieux.

5.7 TEST DU CHI-2

Dans le but de comparer les différentes méthodes d'ajustement, le test du chi-2 introduit à la section 3.3.1 peut être utilisé. Le nombre de classes est égal au maximum de la durée de l'épisode pour chacune des stations et certaines classes sont regroupées ensemble pour que le nombre théorique d'observations dans une classe ne soit pas inférieur à 5. Les tableaux 5.6a et 5.6b contiennent les résultats pour les épisodes pluvieux et secs respectivement.

Les valeurs représentées dans ces deux tableaux indiquent la précision de l'ajustement à un échantillon donné. Voici les principales conclusions tirées de l'analyse des valeurs du chi-2:

- Tout d'abord, la distribution série logarithmique donne de loin les plus mauvais résultats. Cette conclusion était prévisible car les valeurs du paramètre r de la distribution binomiale tronquée sont éloignées de 0. En raison de la mauvaise qualité de l'ajustement la distribution série logarithmique est éliminée pour la suite de l'étude.
- Pour l'ensemble des épisodes, la distribution binomiale négative tronquée donne les meilleurs résultats suivie de près par la distribution géométrique. Cette constatation n'est guère surprenante car la distribution binomiale négative à un paramètre de plus pour ajouter à la précision de l'ajustement.
- En ce qui concerne la meilleure méthode pour estimer les paramètres de la distribution binomiale négative tronquée, les résultats ne sont pas clairs. Il y a seulement une légère tendance en faveur de la méthode des moments dans le cas des épisodes pluvieux. De plus, la règle de calcul des degrés de

liberté du test du chi-2 n'est strictement valable que pour la méthode du maximum de vraisemblance et non pour celle des moments.

Tableau 5.6a Valeurs du chi-2 pour les épisodes pluvieux (GEO, SL, BNT)

Station	GEO	SL	BNT		Nombre de classes
	MV	MV	MV	MM	
QU1	8.67	67.85**	1.17	1.13	8
QU2	17.87**	79.34**	11.75	10.74	8
QU3	5.00	47.98**	4.07	4.12	8
QU4	14.53*	64.24**	8.87	8.37	8
QU5	9.86	48.88**	6.16	6.09	7
NB1	12.70	49.32**	10.66	10.56	8
NB2	9.80	45.94**	9.42	9.16	8
NB3	22.30**	87.43**	10.16*	9.74*	7

* rejet à 5%

** rejet à 1%

Tableau 5.6b Valeurs du chi-2 pour les épisodes secs (GEO, SL, BNT)

Station	GEO	SL	BNT		Nombre de classes
	MV	MV	MV	MM	
QU1	5.48	41.08**	5.70	5.79	10
QU2	9.25	67.15**	10.26	9.15	11
QU3	9.88	40.28**	9.74	9.79	9
QU4	5.52	34.32**	5.47	5.85	10
QU5	9.86	82.70**	9.40	10.80	10
NB1	5.21	58.20**	4.65	4.38	10
NB2	9.05	89.94**	8.50	7.79	13
NB3A	1.39	11.33*	1.43	1.37	7
NB3B	6.09	75.95**	5.95	5.73	12

* rejet à 5%

** rejet à 1%

Il n'est pas approprié de considérer les seules valeurs du chi-2 pour comparer deux distributions n'ayant pas le même nombre de paramètres. La p-value de la valeur du chi-2 définie à l'équation (3.13) permet de juger de la qualité de l'ajustement en tenant compte du nombre de degrés de liberté (donc du nombre de paramètres). Ces résultats sont présentés au tableau 5.7.

Les deux conclusions principales qui ressortent de l'analyse de la p-value

sont:

- Pour les épisodes pluvieux, la distribution binomiale négative tronquée donne les meilleurs résultats pour plusieurs stations;
- Pour les épisodes secs, la distribution géométrique conduit en général aux meilleurs résultats.

Tableau 5.7 P-valeur du chi-2 pour les épisodes pluvieux et secs

Station	Épisode pluvieux			Épisode sec		
	GEO	BNT		GEO	BNT	
	MV	MV	MM	MV	MV	MM
QU1	0.193	0.947	0.951	0.705	0.575	0.564
QU2	0.007**	0.038*	0.057	0.415	0.247	0.330
QU3	0.544	0.539	0.532	0.196	0.136	0.134
QU4	0.024*	0.134	0.137	0.701	0.603	0.557
QU5	0.079	0.188	0.193	0.275	0.225	0.148
NB1	0.048	0.059	0.061	0.735	0.703	0.735
NB2	0.133	0.093	0.103	0.617	0.580	0.649
NB3	0.001**	0.038*	0.045*	-	-	-
NB3A	-	-	-	0.925	0.839	0.849
NB3B	-	-	-	0.808	0.745	0.766

* rejet à 5%

** rejet à 1%

5.8 CRITÈRE D'INFORMATION D'AKAÏKE

Comme on l'a indiqué (section 3.3.3), le critère d'Akaïke permet d'identifier la distribution représentant le mieux les données en tenant compte du nombre de paramètres. Naturellement, la méthode d'estimation des paramètres employée doit être la méthode du maximum de vraisemblance. Le tableau 5.8 contient les valeurs du CIA pour les différentes stations.

La comparaison des deux distributions retenues pour les épisodes pluvieux et secs montre que:

- La distribution binomiale négative tronquée est plus appropriée dans plusieurs cas que la distribution géométrique pour représenter la durée des épisodes

pluvieux;

- La distribution géométrique conduit en général aux valeurs minimums du CIA pour les épisodes secs. Ainsi, l'addition d'un paramètre n'améliore pas suffisamment l'ajustement pour justifier l'utilisation d'une distribution plus complexe (BNT).

Tableau 5.8 Valeurs du CIA pour les épisodes pluvieux et secs

Station	Épisode pluvieux		Épisode sec	
	GEO	BNT	GEO	BNT
QU1	2170.1	2163.5	2473.3	2474.3
QU2	2256.1	2248.7	2719.8	2722.2
QU3	2212.4	2213.8	2272.9	2274.8
QU4	1821.7	1817.2	2191.1	2193.1
QU5	1783.0	1780.5	2349.0	2352.3
NB1	1723.6	1723.7	2135.5	2136.3
NB2	2122.8	2123.5	2778.2	2778.5
NB3	2364.0	2351.4	-	-
NB3A	-	-	680.6	682.7
NB3B	-	-	2659.0	2660.2

Le test du rapport du maximum de vraisemblance (section 3.3.2) permet de déterminer à un niveau de signification α si la distribution binomiale négative tronquée est meilleure que la distribution géométrique. L'équation (3.13) montre le lien entre les valeurs du CIA et la statistique t du test. Soient les hypothèses:

H_0 : les observations proviennent de la loi géométrique (1 paramètre et $r=1$);

H_1 : les observations proviennent de la loi binomiale négative tronquée (2 paramètres).

L'hypothèse H_0 est acceptée à un niveau de signification de 5% si $CIA(GEO) - CIA(BNT) + 2 < 3.84$. Pour ce qui est des épisodes pluvieux, la distribution GEO est rejetée pour 5 stations (NB3, QU1, QU2, QU4 et QU5) sur 8.

En conclusion, mentionnons que si l'on considère l'ensemble des stations les

épisodes secs sont mieux représentés par une distribution géométrique alors que la binomiale négative tronquée est plus appropriée aux épisodes pluvieux.

5.9 LES CYCLES

Un cycle est défini comme étant une séquence composée d'un épisode pluvieux suivi d'un épisode sec (où, inversement, d'un épisode sec suivi d'un épisode pluvieux). Il est alors possible de déterminer la loi de probabilité de la longueur d'un cycle en supposant que les épisodes pluvieux et secs suivent des lois de probabilité données. Dans la perspective de déterminer le couple de distributions représentant mieux les épisodes pluvieux et secs et de porter un regard sur le phénomène du cycle, cette section s'attarde à l'ajustement de la durée des cycles.

5.9.1 Lois de probabilité des cycles (Gabriel et Neumann, 1962)

Dans un premier temps, nous allons déduire les lois de probabilité de la longueur des cycles à partir des lois de probabilité de la durée des épisodes secs et pluvieux.

Soient T_p la variable représentant la durée d'un épisode pluvieux et T_s représentant la durée d'un épisode sec, alors la durée d'un cycle C_{ps} est définie par:

$$C_{ps} = T_p + T_s \quad (5.4)$$

Sous l'hypothèse que T_p et T_s sont indépendants, la probabilité que le cycle soit d'une durée L est donnée par:

$$\begin{aligned} P[C_{ps}=L] &= P[(T_p=1 \text{ et } T_s=L-1) \text{ ou } (T_p=2 \text{ et } T_s=L-2) \dots \text{ou } (T_p=L-1 \text{ et } T_s=1)] \\ &= \sum_{i=0}^{L-2} P[T_p=i+1] \cdot P[T_s=L-i-1] \end{aligned} \quad (5.5)$$

La corrélation entre la durée des épisodes pluvieux qui suit un épisode sec ($T_s T_p$) ou la durée d'un épisode sec qui suit un épisode pluvieux ($T_p T_s$) est très

faible (moins de 0.1 en valeur absolue). L'hypothèse d'indépendance entre ces deux variables est donc acceptée. Cependant, mentionnons que la probabilité d'un jour sec ou pluvieux au début d'un épisode dépend de l'état de fin de l'épisode précédent. Nous ne tenons pas compte de cette dépendance dans notre étude.

Comme on l'a vu précédemment (section 5.7), la distribution géométrique représente le mieux la durée des épisodes secs et la distribution binomiale négative tronquée la durée des épisodes pluvieux. Pour approfondir la comparaison entre ces deux lois, 3 cas sont étudiés pour les cycles:

- a) T_p suit une $GEO(q_p)$ et T_s suit une $GEO(q_s)$ (chaîne de Markov);
- b) T_p suit une $BNT(b_p, r_p)$ et T_s suit une $GEO(q_s)$;
- c) T_p suit une $BNT(b_p, r_p)$ et T_s suit une $BNT(b_s, r_s)$.

Voici les lois de probabilité déduites des ces trois cas:

a) GEO-GEO

Le terme dans la sommation de l'équation 5.5 est

$$P[T_p=i+1]P[T_s=n-i-1] = q_p(1-q_p)^i q_s(1-q_s)^{n-i-2} \quad (5.6a)$$

avec $i=0, \dots, n-2$. La probabilité donne (Gabriel et Neumann, 1962):

$$\begin{aligned} P[C_{ps}=n] &= q_p q_s \sum_{i=0}^{n-2} (1-q_p)^i (1-q_s)^{n-i-2} \\ &= q_s q_p \frac{(1-q_p)^{n-1} - (1-q_s)^{n-1}}{(q_s - q_p)} \end{aligned} \quad (5.6b)$$

b) BNT-GEO

Le terme de la sommation de l'équation 5.5 est:

$$P[T_p=i+1]P[T_s=n-i-1] = \binom{i+r_p}{i+1} b_p^{r_p} \frac{(1-b_p)^{i+1}}{(1-b_p^{r_p})} q_s(1-q_s)^{n-i-2} \quad (5.7a)$$

avec $i=0, \dots, n-2$. La probabilité est donc:

$$P[C_{ps}=n] = \frac{b_p^{r_p} q_s}{(1-b_p^{r_p})} \sum_{i=0}^{n-2} \binom{i+r_p}{i+1} (1-b_p)^{i+1} (1-q_s)^{n-i-2} \quad (5.7b)$$

c) BNT-BNT

Le terme de la sommation de l'équation 5.5 est:

$$P[T_p=i+1]P[T_s=n-i-1] = \binom{i+r_p}{i+1} \frac{b_p^{r_p} (1-b_p)^{i+1}}{(1-b_p^{r_p})} \binom{n-i+r_s-2}{n-i-1} \frac{b_s^{r_s} (1-b_s)^{n-i-1}}{(1-b_s^{r_s})} \quad (5.8a)$$

avec $i=0, \dots, n-2$. La probabilité s'écrit:

$$P[C_{ps}=n] = \frac{b_p^{r_p} b_s^{r_s}}{(1-b_p^{r_p})(1-b_s^{r_s})} \sum_{i=0}^{n-2} \binom{i+r_p}{i+1} \binom{n-i+r_s-2}{n-i-1} (1-b_p)^{i+1} (1-b_s)^{n-i-1} \quad (5.8b)$$

5.9.2 Test du chi-2

Les trois modèles représentant les cycles sont premièrement comparés à l'aide du test du chi-2 (section 3.3.1). Les valeurs des paramètres sont déterminées à partir des valeurs obtenues aux tableaux 5.5a et 5.5b. La méthode d'estimation considérée est celle du maximum de vraisemblance. Le tableau 5.9 contient les valeurs calculées du chi-2 et la p-value correspondante (équation 3.13) pour les différentes stations. Rappelons que la valeur minimum du chi-2 indique le meilleur modèle alors que dans le cas de la p-value c'est la valeur la plus élevée qui indique le meilleur modèle. La station NB3 a été éliminée en raison de l'hétérogénéité mensuelle détectée à la section 5.3.2.

Naturellement, en effectuant la comparaison entre les modèles avec les valeurs du chi-2, le modèle avec le plus de paramètres (BNT-BNT) est le plus performant alors que le modèle avec le moins de paramètres (GEO-GEO) arrive au dernier rang. L'amélioration entre le modèle BNT-GEO et BNT-BNT est cependant faible.

Tableau 5.9 Valeurs du chi-2 pour les cycles

Station	GEO-GEO		BNT-GEO		BNT-BNT		Nombre de classes
	chi-2	p-value	chi-2	p-value	Chi-2	p-value	
NB1	23.9	0.00**	20.9	0.00**	19.1	0.00**	11
NB2	14.8	0.14	13.2	0.16	13.2	0.11	13
QU1	13.3	0.10	11.1	0.13	11.1	0.09	11
QU2	18.1	0.03*	13.6	0.09	13.4	0.06	12
QU3	6.9	0.55	6.8	0.45	6.8	0.34	11
QU4	7.8	0.55	5.6	0.70	5.6	0.59	12
QU5	13.7	0.09	10.3	0.17	9.2	0.16	11

* rejet à 5%

** rejet à 1%

La p-value permet de tenir compte du nombre de degrés de liberté dans le test du chi-2. Pour les stations étudiées, la p-value indique que le modèle le plus performant est BNT-GEO pour les cycles. Ce résultat est conforme avec les résultats obtenus précédemment.

5.9.3 Critère d'information d'Akaike (CIA)

En vue de confirmer les résultats de la section précédente, le critère d'Akaike est appliqué aux 3 modèles. Les résultats se trouvent au tableau 5.10. Il en ressort que le modèle BNT-GEO conduit généralement aux meilleurs résultats. Le modèle GEO-GEO arrive au deuxième rang alors que le modèle BNT-BNT se classe dernier.

Tableau 5.10 Valeurs du critère du CIA pour les cycles

Station	GEO-GEO	Rang	BNT-GEO	Rang	BNT-BNT	Rang
NB1	2484.1	3	2482.8	2	2482.5	1
NB2	3180.5	1	3180.6	2	3181.6	3
QU1	3003.5	2	3002.8	1	3004.6	3
QU2	3225.9	3	3223.1	1	3224.7	2
QU3	2928.3	1	2930.2	2	2932.2	3
QU4	2622.6	2	2621.7	1	2623.8	3
QU5	2719.8	2	2718.6	1	2721.0	3
MOY	2880.7	2.00	2880.0	1.42	2881.5	2.57

Le test du rapport des vraisemblances maximales s'applique aussi ici. Deux cas sont considérés:

a) H_0 : les observations proviennent du modèle GEO-GEO (2 paramètres - $r_p = 1$);
 H_1 : les observations proviennent du modèle BNT-GEO (3 paramètres).

L'hypothèse H_0 est acceptée à un niveau de signification de 5% (équation 3.17) si $CIA(\text{GEO-GEO}) - CIA(\text{BNT-GEO}) + 2 < 3.84$.

b) H_0 : les observations proviennent du modèle GEO-GEO (2 paramètres- $r_p = 1$ et $r_s = 1$);

H_1 : les observations proviennent du modèle BNT-BNT (4 paramètres).

L'hypothèse H_0 est acceptée à un niveau de signification de 5% (équation 3.17) si $CIA(\text{GEO-GEO}) - CIA(\text{BNT-BNT}) + 4 < 5.99$.

Les valeurs du CIA pour les cycles (tableau 5.10) sont utilisées pour effectuer le test du rapport des vraisemblances maximales pour les cas a) et b). Le test indique qu'il y a seulement un cas où le modèle BNT-GEO est meilleur que le modèle GEO-GEO (QU2) et aucun cas où le modèle BNT-BNT est meilleur que le modèle GEO-GEO. Il n'y a pas vraiment contradiction avec les résultats obtenus par le CIA car le test du rapport des vraisemblances maximales à un niveau de signification de 5% est plus sévère que le CIA.

5.10 CONCLUSION

Trois distributions discrètes ont été utilisées pour ajuster la durée des épisodes pluvieux et secs. La région étudiée comprend le Québec (5 stations) et le Nouveau-Brunswick (3 stations). Les hypothèses d'indépendance, d'homogénéité et l'existence de valeurs singulières ont été vérifiées pour les différentes stations. Finalement, les tests utilisés pour comparer les différentes distributions montrent d'une manière générale que:

- La durée des épisodes pluvieux est mieux représentée par une distribution binomiale négative tronquée. La distribution géométrique est quand même adéquate pour représenter ce genre d'échantillon.

- La durée des épisodes secs est mieux représentée par une distribution géométrique. Il n'y a pas d'amélioration apportée par l'ajout d'un paramètre avec la distribution binomiale négative tronquée.

De plus, il a été démontré que l'état (sec ou pluvieux) des jours successifs ne peut pas être considérés comme indépendants.

En terminant, la longueur des cycles (épisode pluvieux et épisode sec) a été également étudiée pour 7 stations (la station NB3 a été éliminée en raison de l'hétérogénéité mensuelle des épisodes secs). Tel que prévu, le modèle considérant une distribution binomiale négative tronquée pour l'épisode pluvieux et une distribution géométrique pour l'épisode sec (BNT-GEO) a conduit aux meilleurs résultats. Le modèle considérant des distributions géométriques pour les épisodes pluvieux et sec (GEO-GEO) a donné quand même des résultats acceptables.

CHAPITRE 6 HAUTEURS DE PRÉCIPITATION

6.1 CHOIX DES STATIONS

Les stations considérées pour l'étude des hauteurs de précipitations sont les mêmes que celles des épisodes pluvieux et secs. Nous référons donc au tableau 5.1 (section 5.1) pour plus de détails sur ces stations. Cependant, rappelons quelques caractéristiques de base des échantillons rencontrés dans ce chapitre:

- les hauteurs de pluie sont mesurées sur une base journalière en mm.
- le seuil de détection est situé à 0.2 mm.
- la période d'enregistrement s'étend du mois de juin au mois de septembre.

6.2 CRITÈRES ET TESTS STATISTIQUES

6.2.1 Indépendance

L'hypothèse d'indépendance des hauteurs de pluie journalière est vérifiée à l'aide du test de Wald-Wolfowitz (section 3.2.1). Les valeurs de la statistique $|u|$ sont calculées à chaque année en raison de l'absence d'observation durant l'hiver. Vu le grand nombre de résultats, seulement les valeurs de $|u|$ dépassant la valeur critique à un niveau de signification de 5% se retrouvent au tableau 6.1. Le coefficient d'autocorrélation d'ordre un (section 3.1.1) est également représenté dans ce tableau.

Tableau 6.1 La statistique $|u|$ du test d'indépendance de Wald-Wolfowitz pour les hauteurs de pluie journalière

Station	Année	$ u $	ρ_1
QU1	1984	2.61	0.31
QU3	1971	2.43	0.28
QU4	1987	2.10	0.21
QU5	1981	2.01	0.24
NB1	1977	1.97	0.21
	1987	3.97	0.50
NB2	1985	2.90	0.36

La figure 6.1 est une représentation graphique "boîtes à moustaches" de l'autocorrélation d'ordre un des hauteurs de pluie journalière calculée à chaque année. Ce genre de graphique permet de juger de l'étalement de ρ_1 (section 5.3.1).

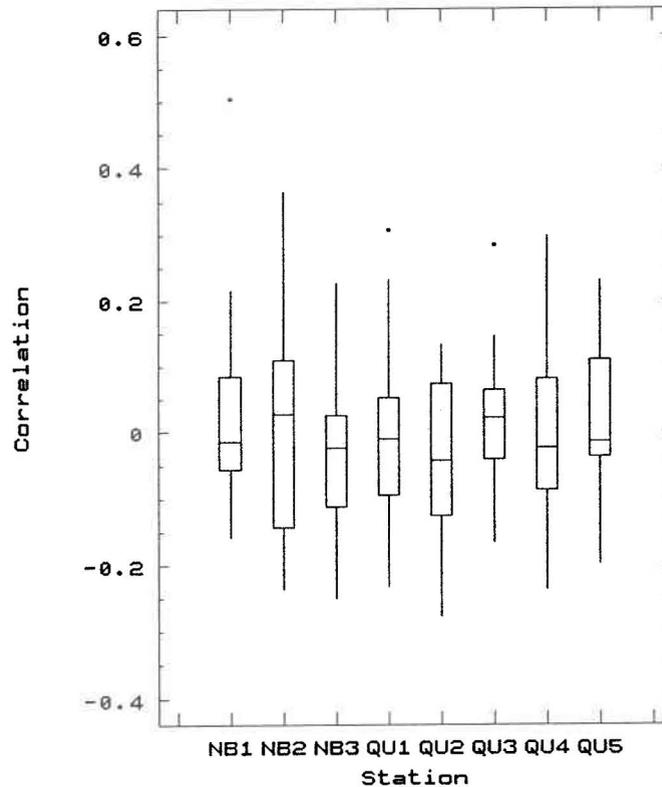


Figure 6.1 Boîtes à moustache de l'autocorrélation d'ordre un des hauteurs de pluie

Comme pour les épisodes pluvieux et secs, il se produit des années où l'indépendance des éléments successifs est discutable. Mais, d'un point de vue statistique, on peut s'attendre à rejeter l'hypothèse d'indépendance alors qu'elle est vraie 12 fois sur 231 (nombre d'années pour toutes les stations) si $\alpha=0.05$ et 2 fois sur 231 si $\alpha=0.01$. Le nombre d'années ayant une dépendance remarquée au tableau 6.1 respecte assez bien ce résultat théorique. De plus, la figure 6.1 montre des médianes voisines de 0 et un étalement réduit. Ainsi, pour les 8 stations étudiées, l'hypothèse d'indépendance des hauteurs de pluie peut être acceptée.

6.2.2 Homogénéité

Pour les raisons mentionnées lors de l'étude des durées des épisodes secs et pluvieux (section 5.3.2), l'homogénéité chronologique n'est pas examinée. L'homogénéité mensuelle, quant à elle, est vérifiée à l'aide du test de Kruskal-Wallis (section 3.2.3). Les hauteurs de pluie sont divisées en quatre sous-échantillons (juin, juillet, août, septembre) selon le mois durant lequel l'enregistrement a été fait et le test est appliqué afin d'identifier la présence de variation saisonnière. Les résultats sont présentés au tableau 6.2.

Tableau 6.2 La statistique τ du test d'homogénéité mensuelle de Kruskal-Wallis pour les hauteurs de pluie

Station	τ
QU1	7.00
QU2	8.72*
QU3	3.36
QU4	0.88
QU5	1.13
NB1	8.60*
NB2	2.00
NB3	3.23

* rejet à 5%

Pour deux stations (QU2 et NB1) on rejette l'hypothèse d'homogénéité à un niveau de signification de 5%. En ce qui concerne les autres stations, l'homogénéité est acceptée à un niveau de 5% avec des valeurs en deçà de la valeur critique. Bien que l'homogénéité ne soit pas rejetée à un niveau de 1%, les échantillons des stations QU2 et NB1 sont quand même divisés en 4 sous-échantillons (juin à septembre). Ces nouveaux sous-échantillons sont désignés par le nom de la station suivi de la lettre M et d'un chiffre désignant le mois.

6.2.3 Valeurs singulières

Le test de Grubbs-Beck (section 3.2.4) est appliqué à chacune des stations. L'hypothèse de log-normalité des observations est acceptée en raison de l'asymétrie positive des échantillons (tableau 6.3). Aucune valeur singulière n'est présente dans

les échantillons étudiées à un niveau de signification de 1%.

6.3 CARACTÉRISTIQUES STATISTIQUES DES HAUTEURS DE PLUIE

Le tableau 6.3 indique la moyenne (\bar{x}) et les coefficients de variation (\hat{C}_v), d'asymétrie (\hat{C}_s) et d'aplatissement (\hat{C}_k) (section 2.1.3b) et la taille (n) des échantillons des hauteurs de pluie journalière.

Tableau 6.3 Caractéristiques statistiques des hauteurs de pluie journalière

Station	\bar{x} (mm)	\hat{C}_v	\hat{C}_s	\hat{C}_k	n
QU1	7.9	1.24	2.48	12.22	1656
QU2M6	6.7	1.23	2.07	7.52	439
QU2M7	6.3	1.32	2.27	8.66	444
QU2M8	7.2	1.37	2.66	12.48	426
QU2M9	8.5	1.38	3.48	20.85	415
QU3	7.3	1.30	3.24	19.93	1715
QU4	6.1	1.34	2.65	12.91	1406
QU5	6.1	1.33	3.02	16.08	1411
NB1M6	6.3	1.23	1.83	6.14	340
NB1M7	7.0	1.36	3.37	19.82	330
NB1M8	7.4	1.21	2.68	13.79	339
NB1M9	6.9	1.51	3.22	15.43	304
NB2	7.0	1.37	2.83	13.99	1618
NB3	8.5	1.45	3.01	16.64	1817
MOY	7.1	1.33	2.77	14.03	

La hauteur moyenne de pluie journalière est d'environ 7 mm. On peut déduire de l'étude des épisodes pluvieux et secs du chapitre précédent qu'il pleut environ 54 jours durant la saison estivale de 122 jours (juin à septembre). Une simple multiplication de ces deux quantités nous donne la hauteur de précipitation moyenne durant l'été: 378 mm.

L'asymétrie des hauteurs de pluie est positive. Son domaine de variation est situé habituellement entre 2 et 3. Le coefficient d'aplatissement, de son côté, est en général très élevé.

La distribution gamma donne un rapport de C_s/C_v égale à 2. L'examen de ce rapport montre que plusieurs stations répondent à ce critère. En particulier, les stations QU1, QU2M9, QU4, NB1M9, NB2 et NB3 ont un quotient C_s/C_v voisin de 2.

La taille des échantillons est d'environ 1500 éléments, exception faite des échantillons mensuels de taille réduite (environ 400).

6.4 AJUSTEMENT DES HAUTEURS DE PLUIE JOURNALIÈRE

6.4.1 Distributions et méthodes d'ajustement considérées

Les histogrammes des hauteurs de pluie mesurées sur une base quotidienne indiquent une diminution de la fréquence des observations à mesure que la hauteur augmente (figure 6.2). Il est souhaitable que les distributions sélectionnées aient cette même propriété. La méthode d'ajustement la plus appropriée pour les grandes tailles d'échantillon est en général celle du maximum de vraisemblance.

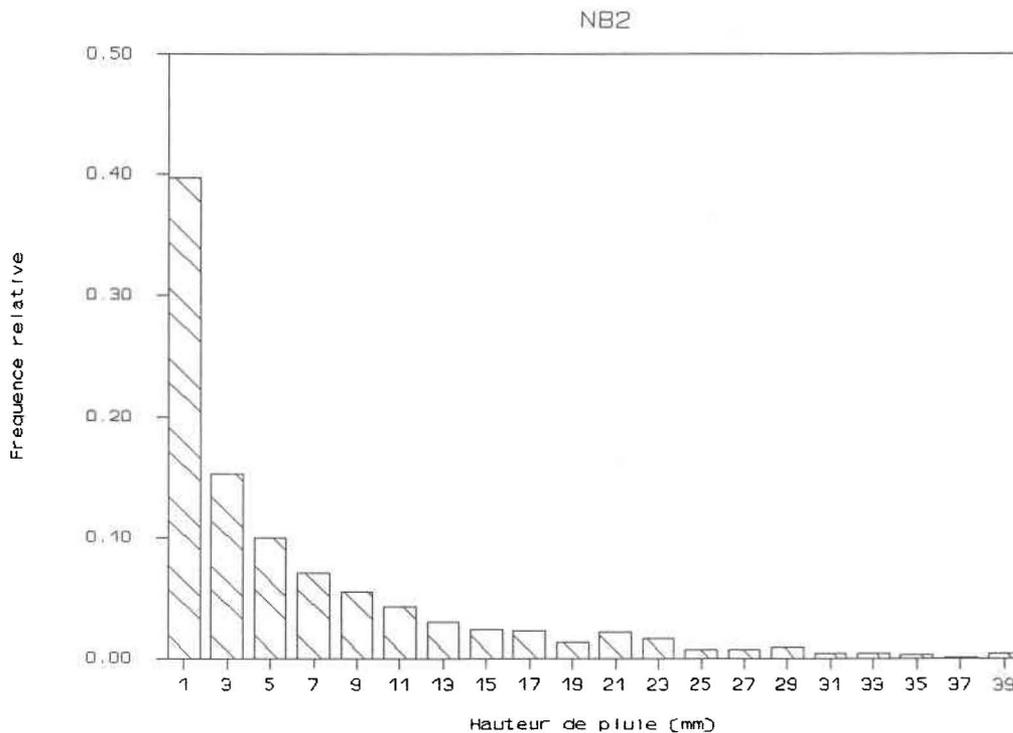


Figure 6.2 Histogramme typique de la hauteur de pluie journalière

Voici la liste des distributions utilisées dans cette étude des hauteurs de pluie journalière:

Exponentielle:	Maximum de vraisemblance
Gamma:	Maximum de vraisemblance
Weibull:	Maximum de vraisemblance
Kappa:	Maximum de vraisemblance
Log-normale:	Maximum de vraisemblance
Exponentielle mixte:	Maximum de vraisemblance
Gamma généralisée:	Maximum de vraisemblance

Les distributions log-normale et gamma généralisée conduisent à une fonction de densité en forme de cloche qui est contradictoire avec la forme des histogrammes. Toutefois, ces deux distributions peuvent avoir un mode très voisin de zéro et peuvent quand même bien représenter les observations.

La distribution Pearson type 3 n'est pas incluse dans la liste ci-dessus car la méthode du maximum de vraisemblance ne donne pas de résultat pour les échantillons considérés. Ceci est dû au fait que $C_s > 2$ (Bobée et Ashkar, 1991).

Les distributions log-gamma et log-Pearson type 3 ne donnent pas de solution pour les échantillons de pluie journalière. La raison vient du fait que la transformation logarithmique des hauteurs de pluie journalière produit des valeurs négatives qui ne sont pas tolérées par le logiciel Ajuste 1.1. Une solution envisagée est de multiplier par une constante les hauteurs de pluie. Cependant, l'ajustement des échantillons transformés sur papier de probabilité n'a pas montré que les lois log-gamma et log-Pearson type 3 sont appropriées pour représenter les hauteurs de pluie.

6.4.2 Calculs effectués

Des programmes informatiques (GAUSS) ont été écrits pour estimer les

paramètres des lois mentionnées dans la section précédente. Aucune erreur ne fut rencontrée durant les ajustements. Le tableau 6.4 contient les paramètres estimés pour chacune des lois.

Rappelons les cas particuliers suivants: la distribution exponentielle est un cas particulier de la distribution gamma lorsque $\lambda=1$, de la distribution Weibull lorsque $c=1$ et de la distribution exponentielle mixte lorsque $p=1$ ou $p=0$; la distribution gamma est un cas particulier de la distribution gamma généralisée lorsque $s=1$. On remarque immédiatement que les distributions gamma et Weibull ont des valeurs de λ et de c inférieures et assez éloignées de un. Il semble donc au premier abord que les distributions gamma et Weibull soient plus appropriées que la loi exponentielle pour représenter les hauteurs de pluie journalière.

Les paramètres de la distribution gamma généralisée sont sujets à de grandes variations. La corrélation entre les paramètres est sans doute à l'origine des fluctuations observées. En effet, il est fréquent d'observer des corrélations élevées entre les paramètres d'une distribution à 3 paramètres.

6.5 TEST DU CHI-2

Le test du chi-2 (section 3.3.1) permet de comparer les ajustements des différentes distributions. Les observations sont divisées en 20 classes équiprobables et, ensuite, le test est appliqué. Les valeurs calculées du chi-2 sont présentées au tableau 6.5.

Une faible valeur du chi-2 indique une bonne adéquation de la loi aux observations. Voici les principaux résultats qui ressortent des valeurs du chi-2:

- La distribution exponentielle mixte représente le mieux les observations.

- Des distributions à deux paramètres, c'est la distribution Weibull qui conduit aux meilleurs résultats. La distribution gamma est légèrement moins

Tableau 6.4 Paramètres estimés pour les hauteurs de pluie journalière

Station	EX	WE		GA		KA		LN		EXM			GG		
	$\hat{\alpha}$	$\hat{\alpha}$	\hat{c}	$\hat{\alpha}$	$\hat{\lambda}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\rho}$	\hat{a}_1	\hat{a}_2	$\hat{\alpha}$	$\hat{\lambda}$	\hat{s}
QU1	0.13	7.1	0.83	0.10	0.77	1.5	4.7	1.3	1.4	0.71	10.4	1.7	7E+00	3.9	1.5
QU2M6	0.15	6.1	0.83	0.12	0.78	1.5	4.0	1.1	1.4	0.67	9.2	1.7	2E+02	7.1	1.5
QU2M7	0.16	5.4	0.78	0.11	0.70	1.3	3.2	1.0	1.4	0.62	9.3	1.4	1E+05	13.0	1.3
QU2M8	0.14	6.1	0.77	0.10	0.69	1.3	3.5	1.1	1.4	0.61	10.9	1.5	2E+06	16.0	1.3
QU2M9	0.12	7.5	0.80	0.09	0.73	1.5	4.9	1.3	1.4	0.73	11.1	1.5	9E+00	4.2	1.5
QU3	0.14	6.7	0.87	0.11	0.84	1.6	4.6	1.3	1.3	0.55	10.9	2.9	1E+07	21.3	1.6
QU4	0.16	5.2	0.78	0.11	0.70	1.3	3.1	1.0	1.4	0.63	9.0	1.3	9E+06	17.8	1.3
QU5	0.16	5.5	0.83	0.13	0.79	1.5	3.6	1.0	1.3	0.54	9.3	2.3	3E+05	15.2	1.5
NB1M6	0.16	5.5	0.81	0.12	0.74	1.3	3.3	1.0	1.4	0.61	9.3	1.5	3E+09	26.5	1.3
NB1M7	0.14	6.3	0.83	0.11	0.78	1.5	4.2	1.2	1.3	0.53	10.9	2.7	1E+05	14.0	1.5
NB1M8	0.14	6.9	0.88	0.11	0.85	1.6	4.9	1.3	1.3	0.68	9.8	2.4	3E+01	5.6	1.6
NB1M9	0.15	5.8	0.77	0.10	0.69	1.3	3.5	1.1	1.4	0.28	15.8	3.5	2E+18	58.5	1.3
NB2	0.14	6.0	0.78	0.10	0.71	1.3	3.6	1.1	1.4	0.59	10.7	1.7	1E+06	15.9	1.3
NB3	0.12	6.8	0.73	0.07	0.63	1.1	3.6	1.2	1.5	0.63	12.7	1.3	1E+12	31.6	1.1
MOY	0.14	6.2	0.81	0.11	0.74	1.4	3.9	1.1	1.4	0.60	10.7	2.0	2E+17	17.9	1.4

performante que la distribution Weibull.

- La distribution exponentielle donne les moins bons résultats en raison de son unique paramètre.

Tableau 6.5 Valeurs du chi-2 pour les hauteurs de pluie journalière

Station	EX	WE	GA	KA	LN	EXM	GG
QU1	194.8**	37.6**	46.5**	102.5**	92.0**	22.4	55.0**
QU2M6	85.2**	33.7**	39.6**	41.2**	39.0**	30.0*	34.0**
QU2M7	103.1**	19.3	25.6	32.8*	26.1	12.3	22.0*
QU2M8	101.1**	28.2*	31.5*	49.8**	42.8**	23.1	37.4**
QU2M9	58.2**	17.9	19.1	37.7**	37.9**	16.9	25.0*
QU3	141.1**	38.6**	57.9**	87.0**	76.6**	30.0*	62.2**
QU4	277.7**	43.3**	58.4**	105.0**	82.0**	19.4	66.3**
QU5	171.6**	33.1*	53.1**	63.3**	54.9**	29.5*	41.5**
NB1M6	81.6**	38.1**	41.2**	55.6**	49.8**	30.0*	45.8**
NB1M7	35.1**	10.9	13.0	25.9	24.7	13.1	19.2*
NB1M8	37.0**	23.0	24.0	42.7**	43.2**	22.7	32.4**
NB1M9	76.8**	27.2	30.9*	37.4**	36.2**	38.2**	34.0**
NB2	287.2**	45.1**	61.5**	108.6**	87.0**	33.1**	69.5**
NB3	577.5**	76.3**	98.1**	123.3**	103.0**	52.0**	95.2**

* rejet à 5%

** rejet à 1%

Plusieurs ajustements sont rejetés à un niveau de signification de 5%. La distribution exponentielle mixte est la plus souvent acceptée à un niveau de signification de 1% et 5%. La distribution Weibull et gamma arrivent respectivement au deuxième et troisième rang. Les autres distributions sont la plupart du temps rejetées à un niveau de signification de 1%.

L'annexe D (figures D.1 à D.6) contient quelques exemples d'ajustement sur papier de probabilité linéaire. Les lois exponentielle, Weibull et exponentielle mixte y sont représentées

6.6 CRITÈRE D'INFORMATION D'AKAÏKE

Le critère d'Akaïke a été introduit à la section 3.3.3. Le CIA permet de tenir compte du nombre de paramètres et de la précision dans l'ajustement. Les valeurs

du CIA sont calculées pour 5 distributions seulement. Les distributions log-normale et gamma généralisée conduisent à des valeurs anormalement basses du CIA (relativement aux autres distributions). Il se peut que la fonction de vraisemblance (L) de ces distribution soit élevée (relativement aux autres distributions) en raison de la réunion des deux facteurs suivants:

- la f.d.p. est telle que $f(0)=0$ pour LN et GG alors qu'elle est telle que $f(0)>0$ pour les autres distributions;
- l'absence d'observations pour des valeurs inférieures à 0.2 mm.

Les distributions log-normale et gamma généralisée sont écartées de la comparaison en raison de l'incertitude entourant les valeurs de leur CIA. Le tableau 6.6 contient les valeurs du CIA pour les lois retenues.

Tableau 6.6 Valeurs du CIA pour les hauteurs de pluie journalière

Station	EX	Rang	WE	Rang	GA	Rang	KA	Rang	EXM	Rang
QU1	10158	5.0	10053	2.0	10078	3.0	10131	4.0	10006	1.0
QU2M6	2556	5.0	2531	2.0	2539	3.0	2546	4.0	2519	1.0
QU2M7	2526	5.0	2474	2.0	2486	3.5	2486	3.5	2454	1.0
QU2M8	2539	5.0	2483	2.0	2496	4.0	2493	3.0	2463	1.0
QU2M9	2611	5.0	2574	2.0	2584	3.0	2588	4.0	2569	1.0
QU3	10257	5.0	10191	2.0	10224	4.0	10185	3.0	10135	1.0
QU4	7913	5.0	7742	2.0	7782	3.0	7783	4.0	7668	1.0
QU5	7911	5.0	7822	2.0	7856	4.0	7828	3.0	7779	1.0
NB1M6	1931	5.0	1904	2.0	1911	3.0	1916	4.0	1884	1.0
NB1M7	1949	5.0	1929	2.0	1937	4.0	1930	3.0	1921	1.0
NB1M8	2037	4.0	2029	2.0	2033	3.0	2038	5.0	2024	1.0
NB1M9	1783	5.0	1740	3.0	1754	4.0	1735	2.0	1731	1.0
NB2	9535	5.0	9342	2.0	9391	4.0	9369	3.0	9271	1.0
NB3	11394	5.0	11026	2.0	11097	4.0	11072	3.0	10910	1.0
MOY	5364	4.9	5274	2.1	5298	3.5	5293	3.3	5238	1.0

La comparaison entre les distributions à partir du CIA mène aux conclusions suivantes:

- La distribution exponentielle mixte conduit aux meilleurs résultats. Ainsi, l'utilisation d'une distribution à 3 paramètres plutôt qu'une à deux paramètres est justifiée.
- Parmi les distributions à deux paramètres, la distribution Weibull représente

le mieux les données. Les distributions Kappa et gamma se retrouvent en troisième et quatrième positions avec des résultats comparables.

- La distribution exponentielle donne les moins bons résultats.

Plusieurs paires de distributions peuvent être comparées à l'aide du test du rapport des vraisemblances maximales. Cependant, les différences entre les valeurs du CIA dans le tableau 6.6 montre clairement que ce test conduit à l'acceptation du modèle avec le plus de paramètres (EXM).

6.7 HAUTEURS DE PRÉCIPITATION ANNUELLE ET MENSUELLE

Cette section traite sommairement de l'ajustement des hauteurs de pluie annuelle et mensuelle. Notre attention est portée plus spécifiquement sur l'identification de la loi la plus adéquate pour ces deux variables de précipitation.

Les échantillons proviennent de deux stations, l'une située au Québec et l'autre au N.-B. Le tableau 6.7 présentent quelques caractéristiques de ces stations. La taille des échantillons est d'environ 80 éléments.

Tableau 6.7 Stations choisies pour l'étude des hauteurs pluie annuelle et mensuelle

Code	Ville	Année d'observation	Moyenne annuelle (mm)
7025280	Montréal	1901-1983	1045
8101600	Frédéricton	1917-1983	1066

L'indépendance des hauteurs de pluie annuelle et mensuelle est en général acceptée à un niveau de signification de 5%. Le test d'indépendance de Wald-Wolfowitz a toutefois détecté un lien à 5% entre les éléments de l'échantillon de Frédéricton pour le mois d'octobre.

Une longue série météorologique peut être influencée par des fluctuations climatiques ou des cycles à long terme. De plus, des changements d'instruments ou d'emplacement de la station ont pu survenir au cours de la période d'enregistrement. Un examen visuel n'a pas permis de détecter d'hétérogénéités évidentes. Une étude plus approfondie des variations météorologiques à long terme est possible avec des tests plus poussés. Cependant, ce genre de tests dépasse la cadre de notre recherche.

Aucune valeur singulière importante ne ressort du test de Grubbs-Beck (à un niveau de 5% et 1%). Le test a été effectué sous l'hypothèse de normalité pour les échantillons ayant un coefficient d'asymétrie près de zéro. La supposition de log-normalité a été appliquée aux échantillons ayant un coefficient d'asymétrie relativement élevé.

Le coefficient d'asymétrie des échantillons de pluie mensuelle et annuelle est positif et est voisin de 0.6. Le coefficient d'aplatissement est habituellement voisin de 4. Le quotient $C_s/C_v \neq 2$ n'indique pas de préférence pour la distribution gamma.

Les distributions considérées pour représenter les hauteurs de pluie annuelle et mensuelle sont les suivantes:

Normale:	Maximum de vraisemblance
Weibull:	Maximum de vraisemblance
Gumbel:	Maximum de vraisemblance
Gamma:	Maximum de vraisemblance
Log-normale:	Maximum de vraisemblance
Gamma Généralisée:	Maximum de vraisemblance
Pearson type 3:	Maximum de vraisemblance
GEV:	Maximum de vraisemblance
Log-normale à 3 paramètres:	Maximum de vraisemblance
Log-Pearson type 3:	Maximum de vraisemblance

La méthode d'estimation des paramètres privilégiée est celle du maximum de vraisemblance car la taille des échantillons est moyenne et la comparaison est faite sur la base du CIA.

Le logiciel AJUSTE 1.1 a été utilisé pour estimer les paramètres de la plupart de lois mentionnées précédemment. Les distributions Pearson type 3 et log-Pearson type 3 n'ont pas donné de solution pour quelques échantillons.

Afin de comparer les distributions entre elles, le CIA est calculé. Les résultats sont inclus au tableau 6.8.

Tableau 6.8 Valeurs du CIA pour les hauteurs de précipitation annuelle et mensuelle

Mois	Montréal									
	N	WE	GU	GA	LN	GG	P3	GEV	LN3	LP3
Jan.	804.9	803.6	814.4	810.9	820.4	805.2	806.9	805.7	806.9	804.5
Fév.	812.1	806.4	802.9	802.3	805.1	804.3	804.3	804.5	804.5	804.3
Mar.	835.2	829.8	829.4	828.6	833.9	830.0	830.2	830.3	830.4	830.0
Avr.	799.7	797.0	795.8	794.4	798.0	796.2	796.2	796.1	796.4	796.1
Mai	850.7	846.8	849.3	853.1	873.2	848.3	848.9	848.9	849.2	849.1
Juin	847.0	842.2	845.3	843.5	850.1	843.7	-	845.2	845.1	844.0
Jui.	844.3	839.4	832.9	832.9	834.9	834.8	834.9	834.6	834.6	835.0
Août	857.0	856.3	858.0	864.7	890.7	857.9	856.2	856.0	856.3	-
Sep.	866.1	860.0	864.4	862.3	869.3	862.0	863.9	864.3	864.7	860.6
Oct.	829.3	827.3	834.9	834.1	846.3	828.6	830.8	829.8	830.9	827.8
Nov.	852.4	843.4	823.0	826.3	822.5	824.4	-	824.3	824.5	824.4
Déc.	828.0	825.8	826.9	824.7	829.3	825.8	826.1	826.1	826.2	825.9
Année	1050.0	1055.0	1058.0	1050.0	1050.7	1052.0	1052.0	1051.0	1051.9	-
Mois	Frédéricton									
	N	WE	GU	GA	LN	GG	P3	GEV	LN3	LP3
Jan.	763.6	759.8	761.6	762.5	772.9	761.6	761.2	761.2	761.1	762.9
Fév.	722.8	718.2	720.6	719.1	724.6	719.8	720.6	720.9	720.9	719.7
Mar.	712.9	710.8	707.6	706.7	708.9	708.7	708.7	708.6	708.7	708.7
Avr.	719.2	718.4	725.5	723.9	733.7	720.4	720.8	719.9	720.8	719.8
Mai	766.1	755.1	751.9	753.4	761.2	755.1	754.5	753.9	754.0	755.9
Juin	751.6	745.2	742.4	741.2	741.4	742.9	741.7	744.3	743.4	742.8
Jui.	743.2	738.3	737.6	736.8	740.3	738.6	738.7	739.0	739.1	738.2
Août	757.3	751.3	754.8	753.1	759.0	753.3	754.9	755.5	755.8	751.0
Sep.	764.2	758.3	757.2	757.6	764.1	759.0	758.8	758.6	758.8	759.2
Nov.	774.0	769.6	770.7	769.3	774.3	770.5	770.9	771.2	771.2	770.6
Déc.	785.3	776.8	777.5	775.5	777.9	777.5	777.1	779.5	778.7	776.4
Année	951.0	960.6	947.4	947.5	946.6	948.2	947.9	948.2	948.1	948.2

En gras: valeur minimale sur la ligne

Une analyse des valeurs du CIA montre de manière générale que:

- Les distributions à 2 paramètres se comportent mieux que les distributions à

3 paramètres, à l'exception des distributions log-normales à 2 et 3 paramètres.

- Les distributions Weibull et gamma conduisent aux meilleurs résultats pour les hauteurs de pluie mensuelle.
- Les distributions log-Pearson type 3 et gamma généralisée donnent les meilleurs résultats dans la catégorie des distributions à 3 paramètres.
- Les distributions normale et log-normale semblent bien représenter les hauteurs de pluie annuelle. Toutefois, l'étude de 2 stations seulement ne nous permet pas d'identifier une distribution plus adéquate qu'une autre pour représenter cette variable.

Le test du rapport de vraisemblance est appliqué à 4 paires de distributions à partir du tableau 6.8. Rappelons que l'hypothèse H_0 est acceptée à un niveau de signification de 5% (équation (3.17)) si $AIC[H_0] - AIC[H_1] + 2 < 3.84$. Voici les résultats obtenus:

a) H_0 : Log-normale LN ($m=0$)

H_1 : Log-normale à 3 paramètres LN3 ($m \neq 0$)

- La distribution log-normale est rejetée dans 15 cas sur 23

b) H_0 : Gamma ($s=1$)

H_1 : Gamma généralisée ($s \neq 1$)

- La distribution gamma est rejetée dans 5 cas sur 23

c) H_0 : Gamma ($m=0$)

H_1 : Pearson type 3 ($m \neq 0$)

- La distribution gamma est rejetée dans 5 cas sur 21.

d) H_0 : Gumbel ($k=0$)

H_1 : GEV ($k \neq 0$)

- La distribution Gumbel est rejetée dans 3 cas sur 23.

Le test du rapport des vraisemblances maximales montre qu'en général une distribution à 2 paramètres est plus appropriée qu'une distribution à 3 paramètres. Comme dans le cas du CIA, la distribution log-normale (LN) ne respecte pas cette règle.

6.8 CONCLUSION

Plusieurs lois statistiques ont été utilisées pour ajuster les hauteurs de précipitation journalière estivale. Les échantillons considérés proviennent de diverses régions du Québec et du N.-B. Les hypothèses d'indépendance et d'homogénéité et la détection de valeurs singulières ont été vérifiées à l'aide de tests appropriés. Les critères du chi-2 et d'Akaike montrent de manière générale que:

- La distribution exponentielle mixte représente le mieux les hauteurs de pluie journalière.
- La distribution Weibull donne les meilleurs résultats parmi les distributions à deux paramètres considérées (gamma, Kappa et log-normale).
- La distribution exponentielle donne les moins bons résultats pour l'ensemble des distributions comparées.

Les hauteurs de précipitation mensuelle de 2 stations ont été également étudiées plus superficiellement. Les valeurs du critère d'Akaike conduisent aux conclusions suivantes:

- Les distributions gamma et Weibull représentent le mieux les observations des

hauteurs de précipitation mensuelle.

- Les distributions plus complexes à trois paramètres n'améliorent pas suffisamment l'ajustement pour remplacer une distribution adéquate à deux paramètres.

CHAPITRE 7 VALEURS EXTRÊMES DE PRÉCIPITATION

7.1 CHOIX DES STATIONS

Le SEA met à la disposition des utilisateurs des tableaux donnant les maximums annuels de pluies pour des durées de 5, 10, 15, 30 minutes, 1, 2, 6, 12 et 24 heures. Les stations disponibles disposent d'au moins sept années d'enregistrement pour lesquelles la période d'enregistrement s'étale du mois d'avril au mois d'octobre.

L'instrument le plus couramment employé est le pluviomètre à augets basculeurs qui fournit des données d'intensité de précipitation pour les courtes durées. C'est un instrument simple et fiable qui a tendance à ne pas mesurer fidèlement les fortes intensités de pluie. Pour cette raison, le SEA rectifie les relevés des pluviomètres à augets basculeurs de manière à ce que les totaux journaliers concordent avec ceux des pluviomètres standards non-enregistreurs. Le pluviomètre à enregistrement numérique automatique longue durée (Fisher & Porter) se retrouve principalement dans les régions isolées. Il est moins répandu et ne donne pas lieu aux mêmes erreurs que le pluviomètre à augets basculeurs pour des taux de précipitation élevés. Cependant il n'est pas efficace pour des durées de moins de 30 minutes.

Pour la région du Québec il y a 107 stations où des séries annuelles de maxima sont disponibles tandis que pour les Provinces Atlantiques il y en a 40. De ces 147 stations, il y en a seulement 16 ayant des relevés portant sur une période de 25 ans ou plus. Plus la période couverte par un relevé est longue, plus l'estimation des paramètres d'une distribution statistique est précise. Pour cette raison, afin de déterminer la loi la plus adéquate pour représenter les séries annuelles de maxima, seulement les stations possédant un relevé minimal de 25 ans et plus ont été retenues. La liste et les principales caractéristiques des stations choisies en raison de leur durée d'enregistrement se trouvent au tableau 7.1.

Tableau 7.1 Stations sélectionnées pour l'étude des séries annuelles de maxima

Nom	Station	Ville	Années d'observation	Années manquantes	Nombre d'années
QU1	7054095	La Pocatière	1962-1986		25 ^b
QU2	7024280	Lennoxville	1960-1986	1963	26 ^b
QU3	7025250	Montréal	1943-1986		44
QU4	7025280	Montréal	1906-1986	1930, 1937, 1956-1958	76 ^a
QU4	7025280	Montréal	1959-1986		28 ^b
QU5	7016280	Québec	1914-1943		30
QU6	7016294	Québec	1961-1986		26
NB1	8100880	Charlo	1959-1986		28
NB2	8101600	Frédéricton	1959-1986		28
NB3	8103200	Moncton	1946-1986	1952, 1957, 1966, 1967	37
NB4	8104900	St. Jean	1958-1986		29
NE1	8204700	Ile de Sable	1962-1986		25
NE2	8205090	Shearwater	1955-1986	1957	31
NE3	8205700	Sydney	1961-1986	1965	25 ^c
NE4	8205990	Truro	1958-1985	1959, 1979	26 ^b
TN1	8401700	Gander	1939-1986	1940, 1943, 1944, 1957	44
TN2	8403606	St. John's	1961-1986	1981	25 ^a
TN2	8403606	St. John's	1961-1986		26 ^b

a durée de 5 min

b durées de 1 h et 24 h

c durées de 5 min et 24 h

Les durées de 5 minutes, 1 et 24 heures (notées 5M, 1H et 24H) ont été retenues car elles représentent la plus petite et la plus grande durée disponible ainsi qu'une durée intermédiaire. Ainsi, nous pourrions examiner si une même distribution peut ajuster adéquatement différentes durées.

Pour alléger le texte et les tableaux, les stations sont désignées par une abréviation composée de deux lettres (la province) suivies d'un chiffre (voir la première colonne tableau 7.1). Par exemple, pour désigner la série annuelle de maxima de Moncton pour une durée de 5 minutes nous écrirons NB3-5M.

7.2 CRITÈRES ET TESTS STATISTIQUES

Avant de passer à l'ajustement, il est essentiel de s'assurer de l'indépendance et de l'homogénéité des éléments des échantillons et de l'absence de valeurs singulières.

7.2.1 Indépendance

Dans le but de vérifier si les éléments des échantillons sont indépendants entre eux, le test de Wald-Wolfowitz décrit à la section 3.2.1 est effectué sur les données retenues. Les résultats du test se trouvent au tableau 7.2.

L'hypothèse d'indépendance de la majorité des stations est acceptée à un niveau de signification de 5%. Cela confirme le fait qu'en général les éléments des séries annuelles de maxima sont indépendants. Il n'y a que deux échantillons qui posent des problèmes d'indépendance. Dans la suite de l'étude, nous ne conservons que les stations pour lesquelles les échantillons sont indépendants, ce qui nous conduit à éliminer les stations QU4-5M et TN1-1H.

Tableau 7.2 La statistique $|u|$ du test d'indépendance de Wald-Wolfowitz pour les séries annuelles de maxima

Station	5 M	1 H	24 H
QU1	-	0.54	1.14
QU2	-	1.58	1.93
QU3	1.61	0.97	0.46
QU4	2.52*	0.68	0.77
QU5	0.89	0.26	0.58
QU6	0.22	1.79	1.61
NB1	0.96	0.09	1.21
NB2	1.26	1.70	0.86
NB3	0.08	1.13	0.85
NB4	0.10	0.22	0.91
NE1	0.15	0.62	0.38
NE2	0.21	1.82	1.46
NE3	1.13	-	1.09
NE4	-	0.55	1.06
TN1	1.24	2.08*	0.12
TN2	0.91	1.00	1.38

* rejet à 5%

7.2.2 Homogénéité

Pour examiner si les éléments d'une station proviennent d'une même population, le test de Mann-Whitney décrit en 3.2.2 est appliqué aux stations.

Pour utiliser le test, il faut au préalable identifier pour chaque station la zone de séparation (coupure) des données en deux sous-échantillons. Habituellement, il y a deux types de division de l'échantillon: par fréquences mensuelles et par périodes chronologiques.

La sélection des deux sous-échantillons par fréquences mensuelles consiste à diviser l'année en deux saisons (par exemple, été et automne), et à classer dans l'une ou l'autre période les observations selon le mois de l'enregistrement. Il est possible que les extrêmes de précipitation d'été n'aient pas la même origine que celles d'automne et proviennent de deux populations différentes (différents phénomènes météorologiques). Les histogrammes des observations selon le mois de l'observation ne permettent pas d'identifier différentes saisons. En effet, les valeurs extrêmes de pluie se produisent principalement durant l'été. De plus, le test

de Mann-Withney appliqué à plusieurs stations n'a pas montré l'évidence d'une séparation des pluies extrêmes en deux saisons différentes. Ainsi, dans cette étude, on considère les valeurs extrêmes comme faisant partie d'une même population.

La sélection des deux sous-échantillons par périodes chronologiques consiste à diviser l'échantillon trié en ordre chronologique selon une coupure fixée à une année spécifique. Les causes de l'hétérogénéité chronologique peuvent être des changements dans l'appareil de lecture ou un déplacement de la station. Les fluctuations climatiques peuvent aussi être source d'hétérogénéité mais sur une courte période d'enregistrement elles sont rarement détectées. Afin de vérifier si des changements ont pu se produire au cours de la période d'enregistrement, un examen visuel des séries chronologiques des valeurs maximums a été effectué pour faire ressortir les années de coupure. Cet examen visuel ne nous a pas permis de mettre en évidence un changement brutale des observations pour les stations considérées. Cependant, par précaution, l'homogénéité est vérifiée dans le cas des interruptions de l'exploitation du pluviomètre ou des changements de position de la station.

Le tableau 7.3 présente les résultats du test de Mann-Whitney sur les stations concernées afin de détecter les hétérogénéités chronologiques. Il est à noter que le test n'est pas effectué pour les stations TN1 à l'année 1940 et NE4 à l'année 1959 car le plus petit sous-échantillon contient moins de trois observations. Cependant, un simple examen de la moyenne des deux sous-échantillons dans le cas des stations TN1 et NE4 ne montre pas d'hétérogénéité évidente.

La taille du plus grand sous-échantillon de la station NB3 ne permet pas d'utiliser l'approximation normale. Dans ce cas, la statistique R^* est calculée à partir de l'équation (3.9). Des tables de valeurs critiques de la statistique R^* de Mann-Withney sont consultées pour interpréter le résultat du test (Rice, 1987). À titre d'exemple, les résultats sont présentés en détail au tableau 7.3 dans le cas de la

station NB3.

Tableau 7.3 La statistique $|u|$ du test d'homogénéité de Mann-Whitney pour les séries annuelles de maxima

Station	Coupure	5M	1H	24H		
QU2	1963	-	0.96	0.92		
NB3	1952	0.12	0.52	0.08		
	1957	0.03	0.08	0.12		
NE3	1965	0.26	-	0.37		
NE4	1979	-	0.49	0.91		
TN1	1943-1944	0.26	0.44	1.12		
	1957	0.35	1.92	0.21		
TN2	1981	0.95	-	-		
		R^*			$R^*_{1\%}$	$R^*_{5\%}$
NB3	1966-1967	331	302	309	258	277

Pour l'ensemble des échantillons considérés, l'hypothèse d'homogénéité est acceptée à un niveau de signification de 5%.

7.2.3 Valeurs singulières

Afin de déterminer si dans un échantillon il y a des valeurs singulières pouvant influencer l'ajustement d'une distribution statistique, le test de Grubbs-Beck décrit à la section 3.2.4 est appliqué. Les résultats du test sont présentés au tableau 7.4.

Tableau 7.4 Résultats du test de détection des valeurs singulières de Grubbs-Beck

Station	Durée	Année	Type
QU2	24H	1984	S*
QU5	5M	1927	S*
QU6	1H	1979	S*
NB2	1H	1976	S*
NB3	1H	1963	S*
NE2	24H	1971	S**
NE3	24H	1981	S*
NE3	5M	1949	S*
TN1	24H	1942	S*

S-supérieure

I-inférieure

* rejet à 5%

** rejet à 1%

Il y a 9 valeurs singulières détectées à un niveau de signification de 5%. Nous obtenons des valeurs singulières supérieures dans tous les cas. Rappelons que le test de Grubbs-Beck n'est utilisé ici qu'à titre indicatif. Lorsque cela fut possible, les valeurs singulières énumérées dans tableau 7.4 ont été validées physiquement par comparaison avec une ou des stations avoisinantes. Cette validation montre qu'il n'y a aucune raison apparente de rejeter les valeurs singulières détectées par le test de Grubbs-Beck.

7.3 CARACTÉRISTIQUES STATISTIQUES DES EXTRÊMES DE PRÉCIPITATION

Avant de procéder aux ajustements des distributions, nous examinons les caractéristiques statistiques des échantillons. Pour chacune des stations épurées après les différents tests de la section précédente, la moyenne (\bar{x}), les coefficients de variation (\hat{C}_v), d'asymétrie (\hat{C}_s) et d'aplatissement (\hat{C}_k) sont calculés. Le tableau 7.5 nous montre les résultats obtenus pour les différentes stations.

On peut remarquer que pour la durée de 5 minutes, la moyenne des maximums annuels varie de 4.8 à 8.7 mm, pour la durée de 1 heure, elle varie de 13.0 à 25.9 mm et pour la durée de 24 heures, elle varie de 45.3 à 85.4 mm. Ces domaines de variation démontrent l'hétérogénéité spatiale des extrêmes de précipitation dans l'est canadien.

Le coefficient d'asymétrie des échantillons est positif sauf pour la station TN2-24H (-0.23) et ne dépasse pas en général 2. Cela indique que la variable du maximum annuel de pluie a une asymétrie vers la droite.

Le coefficient d'aplatissement est habituellement supérieur à 3 et peut varier jusqu'à environs 10 (sauf l'échantillon NE2-24H). Ainsi, les séries annuelles de maxima ont un pic plus accentué que la distribution normale.

Tableau 7.5 Caractéristiques statistiques des séries annuelles de maxima

Station	5 M				1 H				24 H			
	\bar{x} (mm)	\hat{c}_v	\hat{c}_s	\hat{c}_k	\bar{x} (mm)	\hat{c}_v	\hat{c}_s	\hat{c}_k	\bar{x} (mm)	\hat{c}_v	\hat{c}_s	\hat{c}_k
QU1	-	-	-	-	19.6	0.48	1.44	4.79	50.8	0.43	1.66	6.96
QU2	-	-	-	-	22.0	0.30	0.71	3.52	48.6	0.29	1.59	6.82
QU3	8.7	0.31	0.59	3.16	23.6	0.38	1.01	3.59	50.1	0.28	1.04	4.49
QU4	-	-	-	-	25.9	0.33	0.31	3.45	51.8	0.28	0.59	3.07
QU5	8.1	0.35	1.69	7.52	21.7	0.44	1.57	5.73	66.6	0.42	0.92	3.29
QU6	8.0	0.31	1.46	7.75	21.2	0.30	0.51	2.68	59.2	0.19	0.51	2.67
NB1	5.7	0.35	0.76	3.54	16.1	0.34	0.79	3.35	54.8	0.28	0.76	3.68
NB2	7.5	0.36	0.69	4.34	19.0	0.24	1.36	6.67	59.9	0.31	0.95	4.13
NB3	6.5	0.48	1.67	6.50	19.6	0.46	2.39	9.61	57.2	0.40	1.42	5.80
NB4	7.3	0.30	1.25	4.81	24.9	0.35	1.46	5.81	85.4	0.31	1.20	4.47
NE1	7.1	0.39	1.35	5.05	24.4	0.35	1.55	5.01	74.4	0.32	0.93	3.36
NE2	6.4	0.41	0.98	3.41	21.7	0.26	0.21	1.92	69.7	0.31	3.04	15.43
NE3	5.3	0.28	0.19	3.03	-	-	-	-	70.2	0.27	1.78	7.38
NE4	-	-	-	-	19.7	0.33	0.95	3.65	61.4	0.29	0.92	4.50
TN1	4.8	0.50	2.22	10.60	-	-	-	-	45.3	0.35	2.14	8.81
TN2	4.8	0.42	0.97	3.65	17.3	0.28	0.80	4.20	66.6	0.23	-0.23	2.16
MOY	6.7	0.37	1.10	5.11	21.0	0.35	1.02	4.32	69.7	0.31	1.17	5.27

Pour la distribution gamma, le rapport de C_s sur C_v a une valeur théorique de 2. Un examen de ce quotient indique lorsqu'il est voisin de 2 si un échantillon peut être bien représenté par la distribution gamma. Dans notre cas, un faible nombre de stations répond à ce critère (QU3-5M, NB1-5M, NB2-5M, QU4-24H, QU5-24H).

7.4 AJUSTEMENT DES SÉRIES ANNUELLES DE MAXIMA

7.4.1 Distributions et méthodes d'ajustement considérées

Avant de faire un choix définitif des distributions à comparer, plusieurs essais préliminaires d'ajustement ont été analysés sur du papier de probabilité. On observe que 9 distributions peuvent assez bien représenter les séries annuelles de maxima. Les distributions et les méthodes d'estimation retenues sont:

Gumbel	Méthode des moments Maximum de vraisemblance Méthodes des moments pondérés
Gamma	Méthode des moments Maximum de vraisemblance
Log-normale	Méthode des moments Maximum de vraisemblance
Log-gamma	Maximum de vraisemblance sur $\ln(X)$ Méthode des moments sur $\ln(X)$ Méthode des moments sur X
Gamma généralisée	Méthode des moments Maximum de vraisemblance Méthode généralisée des moments SAM Méthode généralisée des moments MM1
Pearson type 3	Méthode des moments Maximum de vraisemblance
GEV	Méthode des moments Maximum de vraisemblance Méthode des moments pondérés
Log-normale à 3	Méthode des moments

paramètres	Maximum de vraisemblance
Log-Pearson type 3	Méthode des moments Maximum de vraisemblance Méthode généralisée des moments SAM Méthode généralisée des moments MM1

Les distributions gamma, log-gamma, gamma généralisée ne sont pas communément employées dans la littérature pour représenter les extrêmes de précipitation alors que les distributions Gumbel et log-normale, quant à elles, sont utilisées intensivement.

7.4.2 Calculs effectués

Le logiciel AJUSTE 1.1 (INRS-Eau,1992) est utilisé pour ajuster toutes les distributions, exception faite des distributions log-normales à 2 et 3 paramètres qui ne sont pas incluses dans ce logiciel.

Une fois les paramètres estimés, la fonction de distribution est déterminée et il est donc possible de calculer la hauteur de pluie correspondant à une probabilité au dépassement p (i.e période de retour $T=1/p$). Le logiciel AJUSTE 1.1 calcule automatiquement les évènements correspondant à 21 probabilités de dépassement. L'intérêt de l'ajustement des séries annuelles de maxima est la hauteur de pluie de temps de retour T plutôt que la valeur des paramètres ajustés qui a une importance limitée. Les probabilités au dépassement retenues dans cette étude sont 0.5, 0.2, 0.1, 0.05 et 0.02 (i.e. période de retour de 2, 5 10, 20 et 50 ans).

L'estimation des paramètres des distributions à deux paramètres n'a causé aucun problème. L'ajustement de ce type de distribution conduit à une solution directe ou encore à un système simple de deux équations à deux inconnues.

En ce qui concerne les distributions à trois paramètres, l'estimation des paramètres a été impossible dans certains cas. Ce type de distribution conduit

souvent à un système complexe de trois équations à trois inconnues nécessitant une méthode itérative (par exemple, Newton-Raphson) pour être résolu. Ces méthodes ne convergent pas toujours ou produisent des erreurs arithmétiques.

Notons plus particulièrement les erreurs suivantes:

- La station TN2-24H provoque plusieurs problèmes. La raison peut être un comportement proche de la loi normale des distributions ajustées ($\hat{C}_s = -0.23$).
- La méthode du maximum de vraisemblance appliquée à la distribution Pearson type 3 ne permet pas d'obtenir de solution pour 9 stations (problèmes de convergence et $|\hat{C}_s^*| > 2$).
- Pour la méthode MM1 appliquée à la distribution log-Pearson type 3, bien que l'algorithme converge la solution est très peu réaliste pour les stations NE2-5M, QU2-1H, NB1-1H, QU4-24H, QU5-24H et NB1-24H. En effet, l'examen visuel sur papier de probabilité montre une courbe théorique très éloignée des observations.

7.5 COMPARAISON ENTRE LES VALEURS OBSERVÉES ET THÉORIQUES

La méthode suivante de comparaison entre les méthodes d'estimation et entre les distributions est basée sur les études de Benson (1968) et de Bobée et Robitaille (1977). Cette méthode empirique consiste à comparer les résultats des hauteurs de pluie calculées à partir des méthodes d'estimation avec les valeurs déduites de l'échantillon pour des périodes de retour fixées.

Afin de calculer les hauteurs de pluie observées correspondant aux différentes périodes de retour, nous avons procédé par interpolation sur papier de probabilité normal. L'interpolation est effectuée de la manière suivante:

- La probabilité empirique correspondant à une probabilité au non-

dépassement P_k est obtenue à partir de la formule de Cunnane (équation 3.20) et est attribuée aux hauteurs de pluie classées en ordre croissant.

- Pour une période de retour T fixée, on détermine la probabilité au dépassement $P=1/T$ et au non-dépassement $P'=1-1/T$ et on en déduit la valeur k telle que $P_k < P' < P_{k+1}$.
- La hauteur de pluie $O(T)$ est alors interpolée entre les valeurs X_k et X_{k+1} et peut être calculé graphiquement sur du papier normal ou directement avec la relation:

$$O(T) = X_k + (u - u_k) \frac{X_k - X_{k+1}}{u_k - u_{k+1}} \quad (7.1)$$

où u , u_k et u_{k+1} correspondent aux variables normales dont les probabilités au non-dépassement sont respectivement P' , P_k et P_{k+1} .

L'interpolation sur papier normal est un choix subjectif et le papier Gumbel peut être aussi utilisé. Des essais montrent qu'un ou l'autre papier de probabilité à peu d'influence sur le résultat de l'interpolation (de l'ordre de 0.2%).

Les résultats de l'interpolation sont présentés dans le tableau 7.6. Les périodes de retour considérées sont 2, 5, 10, 20 et 50 ans (i.e. probabilité au dépassement de 0.50, 0.20, 0.10, 0.05, 0.02). À noter que pour plusieurs stations, la hauteur de pluie $O(50 \text{ ans})$ est obtenue par extrapolation en raison de la faible taille des échantillons. L'extrapolation est basée sur l'équation 7.1 obtenue pour $T=20 \text{ ans}$.

Avec la distribution D ajustée par la méthode d'estimation M , l'écart relatif en pourcentage pour une station donnée est calculé à l'aide de l'équation:

$$E_{D,M}(T) = \frac{H_{D,M}(T) - O(T)}{O(T)} 100 \quad (7.2)$$

Tableau 7.6 Hauteurs de pluie interpolées O(T) entre les observations adjacentes

Station	5M					1H				
	T = 2	T = 5	T = 10	T = 20	T = 50	T = 2	T = 5	T = 10	T = 20	T = 50
QU1	-	-	-	-	-	17.50	24.20	36.41	41.70	45.47*
QU2	-	-	-	-	-	21.65	26.70	32.86	36.91	37.36*
QU3	8.35	11.13	12.45	13.89	15.39	21.00	29.58	39.65	40.92	46.01
QU4	-	-	-	-	-	25.50	33.68	38.05	41.81	46.50
QU5	7.60	9.67	11.37	12.57	17.96	19.45	26.47	34.84	45.58	51.06
QU6	7.80	9.44	10.40	12.21	17.19*	20.60	25.08	28.80	34.67	35.92*
NB1	5.35	7.08	9.11	10.20	10.20*	15.40	20.26	24.75	27.97	28.43*
NB2	7.25	9.39	9.90	13.41	14.79*	18.25	21.80	23.59	27.96	34.94*
NB3	5.80	8.01	10.47	13.75	16.58	16.70	22.49	27.34	43.16	52.17
NB4	6.90	8.10	11.12	12.34	13.64*	24.70	28.28	35.94	48.81	49.83*
NE1	6.40	8.76	10.89	13.89	15.24*	21.70	27.56	40.62	44.56	47.53*
NE2	5.60	8.50	10.87	11.44	12.65	20.10	28.01	29.82	30.33	31.17
NE3	5.30	7.10	7.51	7.86	8.76*	-	-	-	-	-
NE4	-	-	-	-	-	18.90	24.97	31.36	33.57	35.93*
TN1	4.30	6.03	7.73	8.77	12.90	-	-	-	-	-
TN2	4.80	5.98	8.12	8.87	10.49*	16.95	19.24	24.47	27.67	30.03*
	24H									
QU1	48.30	63.27	77.49	99.33	128.00*					
QU2	45.65	58.77	65.65	75.33	98.16*					
QU3	47.90	58.13	70.10	75.68	89.56					
QU4	49.00	60.68	72.68	81.32	84.71					
QU5	59.30	94.75	110.80	126.05	134.05					
QU6	59.25	70.03	78.75	79.48	81.48*					
NB1	53.30	71.82	74.23	81.15	97.96*					
NB2	54.05	73.17	82.54	98.03	112.73*					
NB3	53.10	69.01	84.98	105.05	129.56					
NB4	79.70	107.35	127.88	136.10	161.70*					
NE1	68.60	97.81	114.42	120.28	135.03*					
NE2	67.60	76.59	92.58	115.42	195.92					
NE3	64.20	85.78	93.26	106.73	140.90*					
NE4	59.65	73.57	84.14	95.09	114.48*					
TN1	40.85	51.73	60.91	79.84	101.48					
TN2	68.45	82.70	85.27	87.80	89.43*					

* Obtenue par extrapolation

où $H_{D,M}(T)$ est la hauteur de pluie de période de retour T calculée avec la distribution D et la méthode M pour une station donnée;

$O(T)$ est la hauteur de pluie de période de retour T interpolée entre deux observations adjacentes pour une station donnée.

Afin de comparer les distributions et les méthodes d'estimation des paramètres, la moyenne des écarts en valeurs absolues est calculée:

$$A_{D,M} = \frac{1}{L} \sum |E_{D,M}(T)| \quad (7.3)$$

où L est le nombre de périodes de retour considérées ($L=5$).

Les tableaux 7.7a à 7.7c résument les résultats obtenus pour $A_{D,M}$. Une faible valeur de l'indice empirique $A_{D,M}$ indique un faible écart entre la distribution D ajustée par la méthode M et les points observés.

Il est important de préciser que cette approche permet de comparer les méthodes d'ajustement pour la région de forte probabilité de dépassement et non pour toute la courbe. De plus, elle n'est pas basée sur des critères statistiques.

Il reste maintenant à déterminer la meilleure méthode d'ajustement pour chacune des distributions et, ensuite, identifier la distribution la plus adéquate pour représenter les séries annuelles de maxima.

7.5.1 Comparaison des méthodes d'ajustement

Dans l'optique d'identifier la meilleure méthode d'ajustement pour une distribution D fixée, un classement des valeurs $A_{D,M}$ des méthodes d'ajustement est effectué pour chacune des stations. Le rang 1 correspond à la plus petite valeur de l'indice, le rang 2 à la deuxième plus petite valeur, et ainsi de suite. Lorsque plusieurs méthodes donnent la même valeur de $A_{D,M}$, on attribue le même rang qui correspond à la moyenne des rangs égaux. Cette opération est appliquée à toutes

Tableau 7.7a Moyennes des écarts absolus $A_{D,M}$ (GU, GA, LN et LG)

Station	GU			GA		LN		LG		
	MM	MV	MMP	MM	MV	MM	MV	MVln	MMln	MM
QU3-5M	1.9	2.6	2.0	1.4	1.9	1.4	1.3	3.7	3.4	2.3
QU5-5M	5.5	4.5	5.2	6.8	5.8	5.7	5.0	4.3	4.7	4.8
QU6-5M	6.7	7.4	6.4	7.3	6.7	7.0	6.8	6.9	6.9	6.4
NB1-5M	4.8	4.9	4.9	4.8	4.8	4.8	4.9	6.1	6.0	5.9
NB2-5M	4.7	5.5	4.5	5.7	5.7	4.8	5.2	9.4	7.0	4.7
NB3-5M	7.2	9.3	7.8	8.0	10.0	5.7	7.1	3.3	3.4	5.3
NB4-5M	5.8	6.5	6.0	7.6	8.9	6.4	7.4	6.2	5.6	5.9
NE1-5M	5.0	7.0	5.1	6.2	8.4	4.7	6.2	4.4	3.3	4.3
NE2-5M	5.1	5.5	5.0	4.4	6.4	5.1	5.2	6.4	6.3	6.1
NE3-5M	5.3	8.6	5.9	3.3	3.4	4.4	6.1	11.0	8.9	5.3
TN1-5M	6.4	7.3	5.9	7.3	7.7	5.1	5.3	3.9	3.9	4.2
TN2-5M	5.8	5.3	5.4	6.4	7.2	5.6	4.2	10.8	7.7	5.7
QU1-1H	7.8	10.9	8.3	8.2	11.3	6.5	8.6	7.4	6.1	6.4
QU2-1H	4.5	4.5	4.5	3.8	3.8	4.2	4.2	4.6	4.6	4.4
QU3-1H	4.4	4.2	4.3	4.7	6.1	4.2	4.2	4.1	4.4	4.1
QU4-1H	3.8	5.5	4.2	1.4	1.4	2.5	5.4	8.4	7.4	3.1
QU5-1H	7.1	10.0	7.7	7.9	10.7	6.3	8.6	7.6	6.5	6.1
QU6-1H	3.2	4.6	3.8	2.7	2.8	3.0	3.7	4.1	4.3	2.9
NB1-1H	4.3	4.3	4.0	3.3	3.7	3.9	3.8	4.6	4.5	4.7
NB2-1H	4.2	4.2	4.3	6.1	6.4	5.3	5.6	5.1	5.0	4.9
NB3-1H	14.4	14.7	15.0	15.5	16.7	13.2	14.9	13.8	13.5	12.2
NB4-1H	6.8	7.5	6.9	7.8	8.6	7.0	7.6	7.2	6.7	6.6
NE1-1H	8.4	11.5	9.2	9.9	12.2	8.6	11.0	10.7	9.8	8.4
NE2-1H	7.3	8.2	7.9	6.1	6.0	6.6	7.0	7.5	7.6	6.8
NE4-1H	4.2	4.7	4.0	3.7	5.5	3.8	4.1	4.6	4.4	4.6
TN2-1H	4.0	3.7	3.7	5.0	5.6	4.5	4.2	3.7	3.7	4.5
QU1-24H	6.7	7.6	6.8	7.5	8.1	6.4	6.8	6.2	6.0	6.0
QU2-24H	3.6	5.6	3.9	5.4	6.2	4.4	5.4	5.7	4.8	4.1
QU3-24H	2.1	2.1	2.1	3.7	4.7	2.8	3.3	2.9	2.6	2.6
QU4-24H	2.9	2.9	2.9	3.1	3.3	3.0	3.0	3.0	3.0	3.0
QU5-24H	5.8	6.3	5.5	5.2	6.6	5.9	5.9	6.3	6.1	6.5
QU6-24H	4.4	4.8	4.8	2.9	3.1	3.3	3.3	3.5	3.4	3.5
NB1-24H	3.5	3.6	3.6	3.3	3.3	3.4	3.4	3.5	3.5	3.4
NB2-24H	3.0	3.0	2.8	4.9	5.4	3.7	3.9	3.4	3.1	3.3
NB3-24H	4.9	5.5	5.0	6.4	7.1	4.6	5.1	4.4	4.2	4.2
NB4-24H	3.5	5.4	3.5	4.4	6.5	3.8	5.3	5.5	4.6	4.1
NE1-24H	3.7	4.5	3.6	4.1	5.9	3.5	4.3	4.4	3.8	3.9
NE2-24H	14.5	12.0	11.7	15.2	12.7	14.4	12.1	11.9	11.7	13.7
NE3-24H	4.9	8.0	5.4	6.0	7.6	5.4	7.1	7.6	6.6	5.5
NE4-24H	2.1	2.0	2.0	3.2	3.8	2.6	2.6	2.4	2.2	2.5
TN1-24H	8.7	9.3	9.3	10.4	11.3	9.0	10.1	9.5	9.3	8.4
TN2-24H	8.4	12.9	9.8	6.3	6.7	7.1	8.9	9.2	9.6	7.1

Tableau 7.7b

Moyennes des écarts absolus $A_{D,M}$ (GG, P3 et GEV)

Station	GG				P3		GEV		
	MM	MV	SAM	MM1	MM	MV	MM	MV	MMP
QU3-5M	1.7	1.4	1.2	1.7	1.3	1.0	1.5	1.8	0.9
QU5-5M	4.7	3.7	4.1	4.3	5.2	6.2	4.6	3.9	4.4
QU6-5M	6.4	6.5	-	6.5	6.7	6.6	6.3	6.7	6.4
NB1-5M	4.7	4.7	4.7	4.7	4.7	4.6	4.7	4.7	4.7
NB2-5M	5.7	6.0	6.0	6.0	5.4	5.9	5.6	5.9	6.1
NB3-5M	5.4	5.1	5.1	5.4	4.6	6.9	5.7	5.0	4.2
NB4-5M	5.7	5.6	5.7	5.8	4.9	5.9	5.6	5.6	4.8
NE1-5M	4.5	3.5	3.5	4.5	2.6	-	4.6	3.7	2.8
NE2-5M	4.7	5.5	5.4	-	4.5	4.8	5.3	5.8	6.3
NE3-5M	2.3	2.4	2.4	2.5	2.3	2.5	2.4	2.6	2.3
TN1-5M	4.3	4.2	4.2	4.3	5.2	5.6	4.2	4.0	4.1
TN2-5M	6.1	6.3	6.3	6.4	5.5	6.5	5.7	6.2	5.2
QU1-1H	6.7	-	6.7	6.5	6.5	-	7.0	-	7.6
QU2-1H	3.9	4.3	4.3	4.0	4.2	4.3	3.9	4.2	4.1
QU3-1H	4.2	4.1	4.1	4.2	3.8	3.9	4.2	4.1	4.6
QU4-1H	1.8	2.4	2.3	2.4	1.7	2.3	1.5	2.4	1.3
QU5-1H	6.0	5.5	5.1	6.0	4.5	-	6.2	5.2	4.0
QU6-1H	3.0	2.7	2.8	2.8	2.9	3.9	2.8	2.9	2.9
NB1-1H	3.3	4.6	4.4	3.8	3.8	-	3.5	4.9	4.2
NB2-1H	4.1	4.5	4.3	4.2	3.8	5.1	3.9	4.4	4.3
NB3-1H	11.4	8.5	8.9	10.6	8.0	-	11.4	8.9	7.6
NB4-1H	6.4	6.4	6.3	6.4	6.0	7.2	6.3	6.3	5.7
NE1-1H	7.7	7.4	7.4	7.8	6.2	6.4	7.6	7.6	7.4
NE2-1H	5.4	6.1	6.5	5.8	5.6	7.8	5.7	4.5	5.9
NE4-1H	3.6	5.9	5.0	4.3	4.3	4.8	4.5	6.2	5.0
TN2-1H	4.6	5.0	5.0	4.9	4.1	5.1	4.3	5.0	4.3
QU1-24H	5.9	5.7	5.7	5.8	5.1	6.8	5.9	5.4	4.8
QU2-24H	3.4	2.2	3.5	3.7	3.7	-	3.2	2.6	2.7
QU3-24H	2.3	2.1	2.1	2.3	2.3	2.8	2.1	2.1	1.9
QU4-24H	3.1	3.0	3.0	3.0	2.9	2.7	3.0	2.9	2.9
QU5-24H	5.3	6.4	6.3	5.9	5.2	5.0	6.0	6.9	6.2
QU6-24H	3.2	4.3	3.7	3.5	3.4	-	3.1	4.0	3.8
NB1-24H	3.4	3.5	3.5	-	3.7	4.0	3.5	3.5	3.6
NB2-24H	3.8	2.3	2.9	3.5	2.6	2.2	2.9	2.3	2.3
NB3-24H	4.3	3.3	3.7	4.1	4.0	4.8	4.2	3.0	3.3
NB4-24H	3.5	2.5	3.6	3.9	2.7	2.5	3.5	2.6	2.8
NE1-24H	3.5	5.6	4.2	3.9	3.3	4.2	3.9	6.0	4.9
NE2-24H	-	10.3	10.0	10.2	9.7	-	14.4	10.5	10.5
NE3-24H	4.6	-	4.7	5.0	3.9	4.0	4.4	2.8	3.2
NE4-24H	2.5	2.2	2.3	2.6	2.1	2.4	2.5	2.2	2.2
TN1-24H	6.9	7.8	6.3	6.5	4.8	-	6.7	7.7	4.7
TN2-24H	3.2	-	5.3	4.1	5.7	5.4	3.3	-	2.9

Tableau 7.7c

Moyennes des écarts absolus $A_{D,M}$ (LN3 et LP3)

Station	LN3		LP3			
	MM	MV	BOB	MV	SAM	MM1
QU3-5M	1.5	1.2	1.9	1.5	1.2	1.7
QU5-5M	4.3	4.8	4.7	4.6	4.4	4.4
QU6-5M	6.0	6.4	6.3	6.5	6.5	6.5
NB1-5M	4.7	4.7	4.7	4.7	4.7	4.7
NB2-5M	5.2	5.8	5.9	6.1	6.2	6.1
NB3-5M	4.8	5.7	5.8	5.4	5.2	5.5
NB4-5M	5.3	5.7	6.2	5.7	5.7	5.8
NE1-5M	3.5	4.2	5.1	-	3.4	4.6
NE2-5M	5.1	5.3	4.7	5.4	5.4	-
NE3-5M	2.3	2.5	2.4	2.5	2.4	2.4
TN1-5M	4.2	4.4	4.4	4.2	4.3	4.3
TN2-5M	5.7	6.1	6.5	6.3	6.3	6.4
QU1-1H	6.1	8.1	7.4	9.2	6.7	6.5
QU2-1H	4.2	4.4	3.8	4.3	4.2	-
QU3-1H	4.2	4.1	4.2	4.1	4.1	4.2
QU4-1H	1.8	2.4	2.1	2.1	2.6	2.4
QU5-1H	5.1	4.5	6.6	5.5	4.9	6.0
QU6-1H	2.8	2.8	3.2	2.8	2.8	2.8
NB1-1H	3.9	4.9	3.3	4.9	4.4	-
NB2-1H	3.5	4.7	4.2	4.6	4.3	4.2
NB3-1H	9.6	10.0	11.5	9.4	9.1	10.2
NB4-1H	5.8	6.6	6.6	6.5	6.3	6.4
NE1-1H	7.0	7.0	8.2	7.2	7.1	7.4
NE2-1H	5.6	7.2	5.1	-	6.5	5.8
NE4-1H	4.4	5.9	3.9	6.7	4.9	4.3
TN2-1H	4.2	5.0	5.0	5.0	5.0	4.9
QU1-24H	5.2	6.0	6.2	5.7	5.8	5.9
QU2-24H	3.1	3.1	3.6	-	3.2	3.4
QU3-24H	2.1	2.2	2.6	2.1	2.2	2.3
QU4-24H	3.0	2.9	3.0	3.0	3.0	-
QU5-24H	5.8	6.1	5.3	6.4	6.3	-
QU6-24H	3.4	5.2	3.2	6.2	3.7	3.5
NB1-24H	3.5	3.6	3.4	3.5	3.5	-
NB2-24H	2.8	1.8	4.0	1.8	2.9	3.5
NB3-24H	3.8	3.5	4.5	3.2	3.8	4.1
NB4-24H	3.1	2.5	4.0	2.6	3.4	3.9
NE1-24H	3.7	5.7	4.2	6.4	4.1	3.9
NE2-24H	10.6	11.3	11.8	10.9	11.0	10.9
NE3-24H	3.9	3.4	4.7	3.3	4.2	4.5
NE4-24H	2.0	2.1	-	2.1	2.3	2.6
TN1-24H	5.5	8.7	6.8	8.4	6.3	6.2
TN2-24H	-	-	2.9	-	5.1	3.9

les distributions. De cette manière, les tableaux de rangs 7.8a, 7.8b et 7.8c sont construits à partir des tableaux 7.7a, 7.7b et 7.7c.

Lorsqu'une méthode d'estimation n'a pas donné de solution pour une station, cette dernière est éliminée de la comparaison pour les autres méthodes d'estimation de la distribution. De plus, la méthode MM1 appliquée à la distribution Log-Pearson type 3 est écartée de l'étude en raison de l'incertitude des résultats obtenus (voir section 7.4.2).

Pour chacune des distributions et des méthodes dans les tableaux de rangs, la somme des rangs $S_{D,M}$ est calculée sur l'ensemble des stations considérées pour les durées de 5 minutes, 1 heure et 24 heures et pour l'ensemble des trois durées. Également, le rang $R_{D,M}$ de chacune des méthodes pour une distribution D fixée est déterminé à partir des valeurs de $S_{D,M}$ classées en ordre croissant. L'annexe D (figure D.7 à D.13) apporte un élément visuel à l'ajustement des échantillons en présentant quelques exemples d'ajustement sur papier de probabilité.

L'étude de $S_{D,M}$ et $R_{D,M}$ pour chacune des distributions montre de manière générale que:

Gumbel: Les méthodes des moments et des moments pondérés donnent les meilleurs résultats. La méthode des moments est supérieure à celle des moments pondérés. La méthode du maximum de vraisemblance donne les moins bons résultats.

Gamma: La méthode des moments domine la méthode du maximum de vraisemblance. Ce résultat semble étonnant en tenant compte du fait que les estimateurs du maximum de vraisemblance sont conjointement exhaustifs. Cependant, l'optimalité de la méthode du maximum de vraisemblance ne s'applique pas si les données ne proviennent pas réellement d'une population gamma.

Tableau 7.8a Rangs de la moyenne des écarts absolus (GU, GA, LN et LG)

Station	GU			GA		LN		LG		
	MM	MV	MMP	MM	MV	MM	MV	MVln	MMln	MM
QU3-5M	1.0	3.0	2.0	1.0	2.0	2.0	1.0	3.0	2.0	1.0
QU5-5M	3.0	1.0	2.0	2.0	1.0	2.0	1.0	1.0	2.0	3.0
QU6-5M	2.0	3.0	1.0	2.0	1.0	2.0	1.0	2.5	2.5	1.0
NB1-5M	1.0	2.5	2.5	1.5	1.5	1.0	2.0	3.0	2.0	1.0
NB2-5M	2.0	3.0	1.0	1.5	1.5	1.0	2.0	3.0	2.0	1.0
NB3-5M	1.0	3.0	2.0	1.0	2.0	1.0	2.0	1.0	2.0	3.0
NB4-5M	1.0	3.0	2.0	1.0	2.0	1.0	2.0	3.0	1.0	2.0
NE1-5M	1.0	3.0	2.0	1.0	2.0	1.0	2.0	3.0	1.0	2.0
NE2-5M	2.0	3.0	1.0	1.0	2.0	1.0	2.0	3.0	2.0	1.0
NE3-5M	1.0	3.0	2.0	1.0	2.0	1.0	2.0	3.0	2.0	1.0
TN1-5M	2.0	3.0	1.0	1.0	2.0	1.0	2.0	1.5	1.5	3.0
TN2-5M	3.0	1.0	2.0	1.0	2.0	2.0	1.0	3.0	2.0	1.0
SOMME S _{D,M}	20.0	31.5	20.5	15.0	21.0	16.0	20.0	30.0	22.0	20.0
RANG R _{D,M}	1.0	3.0	2.0	1.0	2.0	1.0	2.0	3.0	2.0	1.0
QU1-1H	1.0	3.0	2.0	1.0	2.0	1.0	2.0	3.0	1.0	2.0
QU2-1H	2.0	2.0	2.0	1.5	1.5	1.5	1.5	2.5	2.5	1.0
QU3-1H	3.0	1.0	2.0	1.0	2.0	1.5	1.5	1.5	3.0	1.5
QU4-1H	1.0	3.0	2.0	1.5	1.5	1.0	2.0	3.0	2.0	1.0
QU5-1H	1.0	3.0	2.0	1.0	2.0	1.0	2.0	3.0	2.0	1.0
QU6-1H	1.0	3.0	2.0	1.0	2.0	1.0	2.0	2.0	3.0	1.0
NB1-1H	2.5	2.5	1.0	1.0	2.0	2.0	1.0	2.0	1.0	3.0
NB2-1H	1.5	1.5	3.0	1.0	2.0	1.0	2.0	3.0	2.0	1.0
NB3-1H	1.0	2.0	3.0	1.0	2.0	1.0	2.0	3.0	2.0	1.0
NB4-1H	1.0	3.0	2.0	1.0	2.0	1.0	2.0	3.0	2.0	1.0
NE1-1H	1.0	3.0	2.0	1.0	2.0	1.0	2.0	3.0	2.0	1.0
NE2-1H	1.0	3.0	2.0	2.0	1.0	1.0	2.0	2.0	3.0	1.0
NE4-1H	2.0	3.0	1.0	1.0	2.0	1.0	2.0	2.5	1.0	2.5
TN2-1H	3.0	1.5	1.5	1.0	2.0	2.0	1.0	1.5	1.5	3.0
SOMME S _{D,M}	22.0	34.5	27.5	16.0	26.0	17.0	25.0	35.0	28.0	21.0
RANG R _{D,M}	1.0	3.0	2.0	1.0	2.0	1.0	2.0	3.0	2.0	1.0
QU1-24H	1.0	3.0	2.0	1.0	2.0	1.0	2.0	3.0	1.5	1.5
QU2-24H	1.0	3.0	2.0	1.0	2.0	1.0	2.0	3.0	2.0	1.0
QU3-24H	2.0	2.0	2.0	1.0	2.0	1.0	2.0	3.0	1.5	1.5
QU4-24H	2.0	2.0	2.0	1.0	2.0	1.5	1.5	2.0	2.0	2.0
QU5-24H	2.0	3.0	1.0	1.0	2.0	1.5	1.5	2.0	1.0	3.0
QU6-24H	1.0	2.5	2.5	1.0	2.0	1.5	1.5	2.5	1.0	2.5
NB1-24H	1.0	2.5	2.5	1.5	1.5	1.5	1.5	2.5	2.5	1.0
NB2-24H	2.5	2.5	1.0	1.0	2.0	1.0	2.0	3.0	1.0	2.0
NB3-24H	1.0	3.0	2.0	1.0	2.0	1.0	2.0	3.0	1.5	1.5
NB4-24H	1.5	3.0	1.5	1.0	2.0	1.0	2.0	3.0	2.0	1.0
NE1-24H	2.0	3.0	1.0	1.0	2.0	1.0	2.0	3.0	1.0	2.0
NE2-24H	3.0	2.0	1.0	2.0	1.0	2.0	1.0	2.0	1.0	3.0
NE3-24H	1.0	3.0	2.0	1.0	2.0	1.0	2.0	3.0	2.0	1.0
NE4-24H	3.0	1.5	1.5	1.0	2.0	1.5	1.5	2.0	1.0	3.0
TN1-24H	1.0	2.5	2.5	1.0	2.0	1.0	2.0	3.0	2.0	1.0
TN2-24H	1.0	3.0	2.0	1.0	2.0	1.0	2.0	2.0	3.0	1.0
SOMME S _{D,M}	26.0	41.5	28.5	17.5	30.5	19.5	28.5	42.0	26.0	28.0
RANG R _{D,M}	1.0	3.0	2.0	1.0	2.0	1.0	2.0	3.0	1.0	2.0
SOMME S _{D,M}	68.0	107.5	76.5	48.5	77.5	52.5	73.5	107.0	76.0	69.0
RANG R _{D,M}	1.0	3.0	2.0	1.0	2.0	1.0	2.0	3.0	2.0	1.0

Tableau 7.8b

Rangs de la moyenne des écarts absolus (GG, P3 et GEV)

Station	GG				P3		GEV		
	MM	MV	SAM	MM1	MM	MV	MM	MV	MMP
QU3-5M	3.5	2.0	1.0	3.5	2.0	1.0	2.0	3.0	1.0
QU5-5M	4.0	1.0	2.0	3.0	1.0	2.0	3.0	1.0	2.0
QU6-5M	-	-	-	-	2.0	1.0	1.0	3.0	2.0
NB1-5M	2.5	2.5	2.5	2.5	2.0	1.0	2.0	2.0	2.0
NB2-5M	1.0	3.0	3.0	3.0	1.0	2.0	1.0	2.0	3.0
NB3-5M	3.5	1.5	1.5	3.5	1.0	2.0	3.0	2.0	1.0
NB4-5M	2.5	1.0	2.5	4.0	1.0	2.0	2.5	2.5	1.0
NE1-5M	3.5	1.5	1.5	3.5	-	-	3.0	2.0	1.0
NE2-5M	-	-	-	-	1.0	2.0	1.0	2.0	3.0
NE3-5M	1.0	2.5	2.5	4.0	1.0	2.0	2.0	3.0	1.0
TN1-5M	3.5	1.5	1.5	3.5	1.0	2.0	3.0	1.0	2.0
TN2-5M	1.0	2.5	2.5	4.0	1.0	2.0	2.0	3.0	1.0
SOMME S _{D,M}	26.0	19.0	20.5	34.5	14.0	19.0	25.5	26.5	20.0
RANG S _{D,M}	3.0	1.0	2.0	4.0	1.0	2.0	2.0	3.0	1.0
QU1-1H	-	-	-	-	-	-	-	-	-
QU2-1H	1.0	3.5	3.5	2.0	1.0	2.0	1.0	3.0	2.0
QU3-1H	3.5	1.5	1.5	3.5	1.0	2.0	2.0	1.0	3.0
QU4-1H	1.0	3.5	2.0	3.5	1.0	2.0	2.0	3.0	1.0
QU5-1H	3.5	2.0	1.0	3.5	-	-	3.0	2.0	1.0
QU6-1H	4.0	1.0	2.5	2.5	1.0	2.0	1.0	2.5	2.5
NB1-1H	1.0	4.0	3.0	2.0	-	-	1.0	3.0	2.0
NB2-1H	1.0	4.0	3.0	2.0	1.0	2.0	1.0	3.0	2.0
NB3-1H	4.0	1.0	2.0	3.0	-	-	3.0	2.0	1.0
NB4-1H	3.0	3.0	1.0	3.0	1.0	2.0	2.5	2.5	1.0
NE1-1H	3.0	1.5	1.5	4.0	1.0	2.0	2.5	2.5	1.0
NE2-1H	1.0	3.0	4.0	2.0	1.0	2.0	2.0	1.0	3.0
NE4-1H	1.0	4.0	3.0	2.0	1.0	2.0	1.0	3.0	2.0
TN2-1H	1.0	3.5	3.5	2.0	1.0	2.0	1.5	3.0	1.5
SOMME S _{D,M}	28.0	35.5	31.5	35.0	10.0	20.0	23.5	31.5	23.0
RANG R _{D,M}	1.0	4.0	2.0	3.0	1.0	2.0	2.0	3.0	1.0
QU1-24H	4.0	1.5	1.5	3.0	1.0	2.0	3.0	2.0	1.0
QU2-24H	2.0	1.0	3.0	4.0	-	-	3.0	1.0	2.0
QU3-24H	3.5	1.5	1.5	3.5	1.0	2.0	2.5	2.5	1.0
QU4-24H	4.0	2.0	2.0	2.0	2.0	1.0	3.0	1.5	1.5
QU5-24H	1.0	4.0	3.0	2.0	2.0	1.0	1.0	3.0	2.0
QU6-24H	1.0	4.0	3.0	2.0	-	-	1.0	3.0	2.0
NB1-24H	-	-	-	-	1.0	2.0	1.5	1.5	3.0
NB2-24H	4.0	1.0	2.0	3.0	2.0	1.0	3.0	1.5	1.5
NB3-24H	4.0	1.0	2.0	3.0	1.0	2.0	3.0	1.0	2.0
NB4-24H	2.0	1.0	3.0	4.0	2.0	1.0	3.0	1.0	2.0
NE1-24H	1.0	4.0	3.0	2.0	1.0	2.0	1.0	3.0	2.0
NE2-24H	-	-	-	-	-	-	1.0	1.5	1.5
NE3-24H	-	-	-	-	1.0	2.0	3.0	1.0	2.0
NE4-24H	3.0	1.0	2.0	4.0	1.0	2.0	3.0	1.5	1.5
TN1-24H	3.0	4.0	1.0	2.0	-	-	2.0	3.0	1.0
TN2-24H	-	-	-	-	2.0	1.0	-	-	-
SOMME S _{D,M}	32.5	26.0	27.0	34.5	17.0	19.0	36.0	28.0	26.0
RANG R _{D,M}	3.0	1.0	2.0	4.0	1.0	2.0	3.0	2.0	1.0
SOMME S _{D,M}	86.5	80.5	79.0	104.0	41.0	58.0	85.0	86.0	69.0
RANG R _{D,M}	3.0	2.0	1.0	4.0	1.0	2.0	2.0	3.0	1.0

Tableau 7.8c

Rangs de la moyenne des écarts absolus (LN3 et LP3)

Station	LN3		LP3		
	MM	MV	BOB	MV	SAM
QU3-5M	2.0	1.0	3.0	2.0	1.0
QU5-5M	1.0	2.0	3.0	2.0	1.0
QU6-5M	1.0	2.0	1.0	2.5	2.5
NB1-5M	1.5	1.5	2.0	2.0	2.0
NB2-5M	1.0	2.0	1.0	2.0	3.0
NB3-5M	1.0	2.0	3.0	2.0	1.0
NB4-5M	1.0	2.0	3.0	1.5	1.5
NE1-5M	1.0	2.0	-	-	-
NE2-5M	1.0	2.0	1.0	2.5	2.5
NE3-5M	1.0	2.0	1.5	3.0	1.5
TN1-5M	1.0	2.0	3.0	1.0	2.0
TN2-5M	1.0	2.0	3.0	1.5	1.5
SOMME $S_{D,M}$	13.5	22.5	24.5	22.0	19.5
RANG $R_{D,M}$	1.0	2.0	3.0	2.0	1.0
QU1-1H	1.0	2.0	2.0	3.0	1.0
QU2-1H	1.0	2.0	1.0	3.0	2.0
QU3-1H	2.0	1.0	3.0	1.5	1.5
QU4-1H	1.0	2.0	1.5	1.5	3.0
QU5-1H	2.0	1.0	3.0	2.0	1.0
QU6-1H	1.5	1.5	3.0	1.5	1.5
NB1-1H	1.0	2.0	1.0	3.0	2.0
NB2-1H	1.0	2.0	1.0	3.0	2.0
NB3-1H	1.0	2.0	3.0	2.0	1.0
NB4-1H	1.0	2.0	3.0	2.0	1.0
NE1-1H	1.5	1.5	3.0	2.0	1.0
NE2-1H	1.0	2.0	-	-	-
NE4-1H	1.0	2.0	1.0	3.0	2.0
TN2-1H	1.0	2.0	2.0	2.0	2.0
SOMME $S_{D,M}$	17.0	25.0	27.5	39.5	21.0
RANG $R_{D,M}$	1.0	2.0	2.0	3.0	1.0
QU1-24H	1.0	2.0	3.0	1.0	2.0
QU2-24H	1.5	1.5	-	-	-
QU3-24H	1.0	2.0	3.0	1.0	2.0
QU4-24H	2.0	1.0	2.0	2.0	2.0
QU5-24H	1.0	2.0	1.0	3.0	2.0
QU6-24H	1.0	2.0	1.0	3.0	2.0
NB1-24H	1.0	2.0	1.0	2.5	2.5
NB2-24H	2.0	1.0	3.0	1.0	2.0
NB3-24H	2.0	1.0	3.0	1.0	2.0
NB4-24H	2.0	1.0	3.0	1.0	2.0
NE1-24H	1.0	2.0	2.0	3.0	1.0
NE2-24H	1.0	2.0	3.0	1.0	2.0
NE3-24H	2.0	1.0	3.0	1.0	2.0
NE4-24H	1.0	2.0	-	-	-
TN1-24H	1.0	2.0	2.0	3.0	1.0
TN2-24H	-	-	-	-	-
SOMME $S_{D,M}$	20.5	24.5	30.0	23.5	24.5
RANG $R_{D,M}$	1.0	2.0	3.0	1.0	2.0
SOMME $S_{D,M}$	51.0	72.0	82.0	75.0	65.0
RANG $R_{D,M}$	1.0	2.0	3.0	2.0	1.0

Log-normale: Ici aussi, la méthode des moments est supérieure à la méthode du maximum de vraisemblance.

Log-gamma: La méthode des moments appliqués à l'échantillon des valeurs observées et celles appliquée à la série transformée donnent les meilleurs résultats. La méthode des moments adaptée à la série originale est supérieure à celle de la série des logarithmes. La méthode du maximum de vraisemblance conduit aux résultats les moins intéressants.

Gamma généralisée: Les méthodes du maximum de vraisemblance et SAM se partagent le premier et le deuxième rang avec des résultats très voisins. La méthode des moments se retrouve au troisième rang alors que la méthode MM1 donne de moins bons résultats. Les résultats n'indiquent pas de tendance remarquable et il est donc difficile de qualifier une méthode drastiquement meilleure qu'une autre.

Pearson type 3: La méthode des moments avec correction d'asymétrie conduit aux meilleurs résultats. De plus, la méthode du maximum de vraisemblance n'offre pas de solution pour plusieurs stations.

GEV: La méthode des moments pondérés mène aux meilleurs résultats. Les méthodes des moments et du maximum de vraisemblance se classent deuxième et troisième avec des valeurs de $S_{GEV,M}$ assez rapprochées.

Log-normale à 3 paramètres: La méthode des moments conduit à de meilleurs résultats que la méthode du maximum de vraisemblance.

Log-Pearson type 3: la méthode SAM conduit aux meilleurs résultats. La méthode du maximum de vraisemblance se retrouve au deuxième rang alors que la méthode BOB se situe au troisième rang. À noter que la méthode

MM1 aurait donné de bons résultats si elle avait été incluse dans la comparaison. Il est important de mentionner que les quatre méthodes présentent des valeurs de $A_{LP3,M}$ voisins.

Il est maintenant possible de déterminer la méthode d'ajustement pour chacune des distributions qui serviront à la comparaison de ces dernières. La sélection finale est effectuée sur la base du $S_{D,M}$. La capacité de la méthode de produire une solution est également considérée lorsqu'il y a peu de différence entre les valeurs de $S_{D,M}$. Voici les méthodes retenues pour la suite de l'étude:

Gumbel	méthode des moments
Gamma	méthode des moments
Log-normale	méthode des moments
Log-gamma	méthode des moments sur X
Gamma généralisée	SAM
Pearson type 3	méthode des moments
GEV	méthode des moments pondérés
Log-normale à 3 paramètres	méthode de moments
Log-Pearson type 3	SAM

Les meilleures méthodes d'estimation pour les distributions à 3 paramètres ne sont pas ressorties clairement. Plus particulièrement, les méthodes d'estimation pour les distributions gamma généralisée et log-Pearson type 3 nécessiteraient une étude approfondie afin d'identifier la méthode d'estimation la plus adéquate dans chacun des cas.

7.5.2 Comparaison des distributions

Les valeurs des $A_{D,M}$ de la méthode la plus efficace pour chacune des distributions sont présentées au tableau 7.9. D'une manière similaire à la section précédente, le tableau de rangs 7.10 est construit à partir du tableau 7.9 afin de comparer d'une manière objective les différentes distributions.

Tableau 7.9 Moyennes des écarts absolus A_D pour les méthodes sélectionnées

Station	GU	GA	LN	LG	GG	P3	GEV	LN3	LP3
	MM	MM	MM	MM	SAM	MM	MMP	MM	SAM
QU3-5M	1.9	1.4	1.4	2.3	1.2	1.3	0.9	1.5	1.2
QU5-5M	5.5	6.8	5.7	4.8	4.1	5.2	4.4	4.3	4.4
QU6-5M	6.7	7.3	7.0	6.4	-	6.7	6.4	6.0	6.5
NB1-5M	4.8	4.8	4.8	5.9	4.7	4.7	4.7	4.7	4.7
NB2-5M	4.7	5.7	4.8	4.7	6.0	5.4	6.1	5.2	6.2
NB3-5M	7.2	8.0	5.7	5.3	5.1	4.6	4.2	4.8	5.2
NB4-5M	5.8	7.6	6.4	5.9	5.7	4.9	4.8	5.3	5.7
NE1-5M	5.0	6.2	4.7	4.3	3.5	2.6	2.8	3.5	3.4
NE2-5M	5.1	4.4	5.1	6.1	5.4	4.5	6.3	5.1	5.4
NE3-5M	5.3	3.3	4.4	5.3	2.4	2.3	2.3	2.3	2.4
TN1-5M	6.4	7.3	5.1	4.2	4.2	5.2	4.1	4.2	4.3
TN2-5M	5.8	6.4	5.6	5.7	6.3	5.5	5.2	5.7	6.3
QU1-1H	7.8	8.2	6.5	6.4	6.7	6.5	7.6	6.1	6.7
QU2-1H	4.5	3.8	4.2	4.4	4.3	4.2	4.1	4.2	4.2
QU3-1H	4.4	4.7	4.2	4.1	4.1	3.8	4.6	4.2	4.1
QU4-1H	3.8	1.4	2.5	3.1	2.3	1.7	1.3	1.8	2.6
QU5-1H	7.1	7.9	6.3	6.1	5.1	4.5	4.0	5.1	4.9
QU6-1H	3.2	2.7	3.0	2.9	2.8	2.9	2.9	2.8	2.8
NB1-1H	4.3	3.3	3.9	4.7	4.4	3.8	4.2	3.9	4.4
NB2-1H	4.2	6.1	5.3	4.9	4.3	3.8	4.3	3.5	4.3
NB3-1H	14.4	15.5	13.2	12.2	8.9	8.0	7.6	9.6	9.1
NB4-1H	6.8	7.8	7.0	6.6	6.3	6.0	5.7	5.8	6.3
NE1-1H	8.4	9.9	8.6	8.4	7.4	6.2	7.4	7.0	7.1
NE2-1H	7.3	6.1	6.6	6.8	6.5	5.6	5.9	5.6	6.5
NE4-1H	4.2	3.7	3.8	4.6	5.0	4.3	5.0	4.4	4.9
TN2-1H	4.0	5.0	4.5	4.5	5.0	4.1	4.3	4.2	5.0
QU1-24H	6.7	7.5	6.4	6.0	5.7	5.1	4.8	5.2	5.8
QU2-24H	3.6	5.4	4.4	4.1	3.5	3.7	2.7	3.1	3.2
QU3-24H	2.1	3.7	2.8	2.6	2.1	2.3	1.9	2.1	2.2
QU4-24H	2.9	3.1	3.0	3.0	3.0	2.9	2.9	3.0	3.0
QU5-24H	5.8	5.2	5.9	6.5	6.3	5.2	6.2	5.8	6.3
QU6-24H	4.4	2.9	3.3	3.5	3.7	3.4	3.8	3.4	3.7
NB1-24H	3.5	3.3	3.4	3.4	3.5	3.7	3.6	3.5	3.5
NB2-24H	3.0	4.9	3.7	3.3	2.9	2.6	2.3	2.8	2.9
NB3-24H	4.9	6.4	4.6	4.2	3.7	4.0	3.3	3.8	3.8
NB4-24H	3.5	4.4	3.8	4.1	3.6	2.7	2.8	3.1	3.4
NE1-24H	3.7	4.1	3.5	3.9	4.2	3.3	4.9	3.7	4.1
NE2-24H	14.5	15.2	14.4	13.7	10.0	9.7	10.5	10.6	11.0
NE3-24H	4.9	6.0	5.4	5.5	4.7	3.9	3.2	3.9	4.2
NE4-24H	2.1	3.2	2.6	2.5	2.3	2.1	2.2	2.0	2.3
TN1-24H	8.7	10.4	9.0	8.4	6.3	4.8	4.7	5.5	6.3
TN2-24H	8.4	6.3	7.1	7.1	5.3	5.7	2.9	-	5.1

Tableau 7.10

Rangs de la moyenne des écarts absolus A_D pour les méthodes sélectionnées

Station	GU	GA	LN	LG	GG	P3	GEV	LN3	LP3
	MM	MM	MM	MM	SAM	MM	MMP	MM	SAM
QU3-5M	8.0	5.5	5.5	9.0	2.5	4.0	1.0	7.0	2.5
QU5-5M	7.0	9.0	8.0	5.0	1.0	6.0	3.5	2.0	3.5
QU6-5M	-	-	-	-	-	-	-	-	-
NB1-5M	7.0	7.0	7.0	9.0	3.0	3.0	3.0	3.0	3.0
NB2-5M	1.5	6.0	3.0	1.5	7.0	5.0	8.0	4.0	9.0
NB3-5M	8.0	9.0	7.0	6.0	4.0	2.0	1.0	3.0	5.0
NB4-5M	6.0	9.0	8.0	7.0	4.5	2.0	1.0	3.0	4.5
NE1-5M	8.0	9.0	7.0	6.0	4.5	1.0	2.0	4.5	3.0
NE2-5M	4.0	1.0	4.0	8.0	6.5	2.0	9.0	4.0	6.5
NE3-5M	8.5	6.0	7.0	8.5	4.5	2.0	2.0	2.0	4.5
TN1-5M	8.0	9.0	6.0	3.0	3.0	7.0	1.0	3.0	5.0
TN2-5M	6.0	9.0	3.0	4.5	7.5	2.0	1.0	4.5	7.5
SOMME S_D	72.0	79.5	65.5	67.5	48.0	36.0	32.5	40.0	54.0
RANG R_D	8.0	9.0	6.0	7.0	4.0	2.0	1.0	3.0	5.0
QU1-1H	8.0	9.0	3.5	2.0	5.5	3.5	7.0	1.0	5.5
QU2-1H	9.0	1.0	4.5	8.0	7.0	4.5	2.0	4.5	4.5
QU3-1H	7.0	9.0	5.5	3.0	3.0	1.0	8.0	5.5	3.0
QU4-1H	9.0	2.0	6.0	8.0	5.0	3.0	1.0	4.0	7.0
QU5-1H	8.0	9.0	7.0	6.0	4.5	2.0	1.0	4.5	3.0
QU6-1H	9.0	1.0	8.0	6.0	3.0	6.0	6.0	3.0	3.0
NB1-1H	6.0	1.0	3.5	9.0	7.5	2.0	5.0	3.5	7.5
NB2-1H	3.0	9.0	8.0	7.0	5.0	2.0	5.0	1.0	5.0
NB3-1H	8.0	9.0	7.0	6.0	3.0	2.0	1.0	5.0	4.0
NB4-1H	7.0	9.0	8.0	6.0	4.5	3.0	1.0	2.0	4.5
NE1-1H	6.5	9.0	8.0	6.5	4.5	1.0	4.5	2.0	3.0
NE2-1H	9.0	4.0	7.0	8.0	5.5	1.5	3.0	1.5	5.5
NE4-1H	3.0	1.0	2.0	6.0	8.5	4.0	8.5	5.0	7.0
TN2-1H	1.0	8.0	5.5	5.5	8.0	2.0	4.0	3.0	8.0
SOMME S_D	93.5	81.0	83.5	87.0	74.5	37.5	57.0	45.5	70.5
RANG R_D	9.0	6.0	7.0	8.0	5.0	1.0	3.0	2.0	4.0
QU1-24H	8.0	9.0	7.0	6.0	4.0	2.0	1.0	3.0	5.0
QU2-24H	5.0	9.0	8.0	7.0	4.0	6.0	1.0	2.0	3.0
QU3-24H	3.0	9.0	8.0	7.0	3.0	6.0	1.0	3.0	5.0
QU4-24H	2.0	9.0	6.0	6.0	6.0	2.0	2.0	6.0	6.0
QU5-24H	3.5	1.5	5.0	9.0	7.5	1.5	6.0	3.5	7.5
QU6-24H	9.0	1.0	2.0	5.0	6.5	3.5	8.0	3.5	6.5
NB1-24H	5.5	1.0	2.5	2.5	5.5	9.0	8.0	5.5	5.5
NB2-24H	6.0	9.0	8.0	7.0	4.5	2.0	1.0	3.0	4.5
NB3-24H	8.0	9.0	7.0	6.0	2.0	5.0	1.0	3.5	3.5
NB4-24H	5.0	9.0	7.0	8.0	6.0	1.0	2.0	3.0	4.0
NE1-24H	3.5	6.5	2.0	5.0	8.0	1.0	9.0	3.5	6.5
NE2-24H	8.0	9.0	7.0	6.0	2.0	1.0	3.0	4.0	5.0
NE3-24H	6.0	9.0	7.0	8.0	5.0	2.5	1.0	2.5	4.0
NE4-24H	2.5	9.0	8.0	7.0	5.5	2.5	4.0	1.0	5.5
TN1-24H	7.0	9.0	8.0	6.0	4.5	2.0	1.0	3.0	4.5
TN2-24H	-	-	-	-	-	-	-	-	-
SOMME S_D	82.0	109.0	92.5	95.5	74.0	47.0	49.0	50.0	76.0
RANG R_D	6.0	9.0	7.0	8.0	4.0	1.0	2.0	3.0	5.0
SOMME S_D	247.5	269.5	241.5	250.0	196.5	120.0	138.5	135.5	200.5
RANG R_D	7.0	9.0	6.0	8.0	4.0	1.0	3.0	2.0	5.0

L'examen des rangs R_D et la somme des rangs S_D indique clairement une supériorité des distributions à trois paramètres face aux distributions à deux paramètres. Il est normal qu'en général une distribution à trois paramètres ajuste mieux les observations qu'une distribution à deux paramètres. En effet, le paramètre supplémentaire ajoute à la précision de l'ajustement. Pour cette raison, nous allons traiter séparément les distributions à deux paramètres et celles à trois paramètres.

Distribution à trois paramètres

Les distributions Pearson type 3, log-normale à 3 paramètres et GEV conduisent aux meilleurs résultats. Les distributions Gamma généralisée et Log-Pearson type 3, quant à eux, se situent moins bien. Bien que la distribution Pearson type 3 est supérieure aux autres, il est difficile d'attribuer le premier rang à une des distributions plutôt qu'à une autre. Il est plus réaliste de dire que toutes ces distributions sont en générale adéquates à l'ajustement des séries annuelles de maxima.

Distribution à deux paramètres

La distribution log-normale conduit en général aux meilleurs résultats. La distribution Gumbel se comporte très bien pour la durée de 24 heures et donne de moins bons résultats pour les durées de 5 minutes et 1 heure. La distribution log-gamma se situe au troisième rang. La distribution Gamma conduit au moins bons résultats sauf pour la durée de 1 heure. Pour l'ensemble des trois durées les distributions log-normale, Gumbel et log-gamma donnent des résultats voisins alors que la loi gamma donne les moins bons résultats.

7.6 CRITÈRE D'INFORMATION D'AKAÏKE

Dans le cas des séries annuelles de maxima, il est essentiel de choisir une distribution avec le moins de paramètres possibles afin de diminuer la variance des pluies estimées de période de retour T . La comparaison à l'aide de la moyenne des écarts en valeurs absolues $A_{D,M}$ ne permet pas savoir s'il y a une amélioration

significative dans l'ajustement en ayant une distribution avec un paramètre de plus. Le critère d'Akaike (section 3.3.3) rend possible la quantification de cette amélioration. Le tableau 7.11 indique les valeurs du CIA pour les différentes distributions et les différentes stations. La valeur minimum du CIA indique la meilleure distribution pour une station donnée. Il est important de mentionner que cette comparaison entre les distributions est possible lorsque les paramètres sont estimés par la méthode du maximum de vraisemblance.

Comme à la section 7.5.1, le tableau de rangs 7.12 est construit à partir du tableau 7.11. La méthode du maximum de vraisemblance appliquée à la distribution Pearson type 3 est éliminée de la comparaison à cause de son incapacité à donner des solutions pour plusieurs stations.

La comparaison des distributions montre que:

- De manière générale, l'utilisation d'une distribution à trois paramètres n'améliore pas suffisamment l'ajustement pour la substituer à une distribution à deux paramètres appropriée.
- Dans le cas des distributions à deux paramètres, la distribution log-gamma conduit aux meilleurs résultats sauf pour la durée de 5 minutes. Les distributions Gumbel et log-normale se partagent la deuxième et la troisième position. La distribution gamma conduit aux moins bons résultats.
- Pour les distributions à trois paramètres, la distribution log-normale à 3 paramètres se classe au premier rang. Dans la grande majorité des cas, les valeurs du CIA sont similaires pour les distributions à trois paramètres et, pour cette raison, il est difficile de distinguer nettement une de ces distributions.

Tableau 7.11

Valeurs du CIA pour les séries annuelles de maxima

Station	GU	GA	LN	LG	GG	P3	GEV	LN3	LP3
QU3-5M	212.8	212.3	212.4	213.5	214.2	214.0	214.2	214.3	214.2
QU5-5M	138.9	141.9	139.6	138.1	139.2	136.6	139.2	138.2	138.3
QU6-5M	119.7	120.0	119.6	120.5	121.6	121.9	121.6	121.5	121.6
NB1-5M	119.2	119.1	119.1	120.2	121.0	120.8	121.1	121.0	121.0
NB2-5M	136.5	135.7	136.8	139.4	137.6	137.6	137.7	137.6	137.7
NB3-5M	178.4	180.1	177.5	177.5	179.0	179.7	179.0	179.1	179.1
NB4-5M	125.1	127.1	125.7	125.0	126.9	126.9	126.9	126.9	126.9
NE1-5M	118.0	119.6	117.8	116.6	118.4	-	118.4	117.3	-
NE2-5M	142.1	145.0	142.0	142.1	143.8	144.0	143.7	143.8	143.8
NE3-5M	97.5	95.9	97.1	100.2	97.4	97.5	97.2	97.5	97.2
TN1-5M	184.0	186.5	182.5	181.3	183.2	182.6	183.4	183.0	183.2
TN2-5M	104.7	104.8	104.8	107.9	106.6	106.6	106.7	106.6	106.6
QU1-1H	176.3	177.7	175.2	173.9	174.7	-	-	173.7	177.2
QU2-1H	173.4	173.5	173.2	173.3	175.2	174.8	175.3	175.2	175.2
QU3-1H	311.8	313.4	311.7	311.3	313.3	313.1	313.4	313.2	313.3
QU4-1H	204.7	203.3	205.4	207.5	204.6	204.6	204.5	204.6	204.6
QU5-1H	212.0	214.2	211.4	210.0	210.6	-	210.6	209.3	209.5
QU6-1H	165.0	164.3	164.5	164.9	166.3	166.1	166.4	166.4	166.3
NB1-1H	172.5	173.1	172.4	172.2	174.3	-	174.4	174.0	174.2
NB2-1H	163.0	164.5	163.5	163.1	164.9	165.2	164.9	165.0	164.9
NB3-1H	247.8	253.9	248.1	245.4	243.7	-	243.7	243.9	243.7
NB4-1H	202.5	204.6	202.9	202.4	204.2	204.5	204.2	204.3	204.3
NE1-1H	169.9	174.0	171.5	170.2	168.5	167.9	168.5	168.3	168.2
NE2-1H	197.6	197.0	197.1	197.2	199.0	198.1	198.7	199.1	-
NE4-1H	168.8	170.2	169.0	168.5	170.2	168.2	170.2	169.5	169.7
TN2-1H	156.7	156.5	156.4	156.6	158.3	158.4	158.2	158.3	158.3
QU1-24H	220.1	221.4	219.8	219.4	221.4	221.4	221.4	221.4	221.4
QU2-24H	205.1	208.4	206.6	205.8	205.8	-	205.8	204.8	-
QU3-24H	351.3	353.3	351.9	351.4	353.3	353.5	353.2	353.3	353.3
QU4-24H	229.4	229.6	229.3	229.2	231.2	230.4	231.3	231.1	231.2
QU5-24H	281.8	282.4	281.1	280.8	282.8	281.4	283.0	282.3	282.6
QU6-24H	199.4	200.2	199.8	199.7	201.4	-	201.3	201.0	201.0
NB1-24H	232.9	233.4	232.9	232.8	234.8	234.0	234.9	234.6	234.8
NB2-24H	240.9	242.1	241.1	240.8	242.7	241.8	242.8	242.4	242.6
NB3-24H	329.0	330.9	328.7	328.0	329.8	329.3	330.0	329.6	329.7
NB4-24H	266.9	269.5	267.8	267.3	268.1	267.3	268.1	267.8	267.9
NE1-24H	227.5	229.0	227.9	227.4	228.7	227.3	228.8	228.1	228.3
NE2-24H	280.4	286.0	281.4	279.9	279.4	-	279.5	279.7	279.7
NE3-24H	211.2	215.8	213.7	212.9	210.4	209.3	210.5	209.9	209.7
NE4-24H	222.3	222.9	222.3	222.2	224.2	223.8	224.3	224.1	224.2
TN1-24H	349.7	356.5	351.9	350.4	350.8	-	350.4	352.0	351.4
TN2-24H	222.5	220.0	221.0	221.6	-	221.5	-	-	-

Tableau 7.12

Rangs de la valeur du CIA

Station	GU	GA	LN	LG	GG	GEV	LN3	LP3
QU3-5M	3.0	1.0	2.0	4.0	6.0	6.0	8.0	6.0
QU5-5M	4.0	8.0	7.0	1.0	5.5	5.5	2.0	3.0
QU6-5M	2.0	3.0	1.0	4.0	7.0	7.0	5.0	7.0
NB1-5M	3.0	1.5	1.5	4.0	6.0	8.0	6.0	6.0
NB2-5M	2.0	1.0	3.0	8.0	4.5	6.5	4.5	6.5
NB3-5M	3.0	8.0	1.5	1.5	4.5	4.5	6.5	6.5
NB4-5M	2.0	8.0	3.0	1.0	5.5	5.5	5.5	5.5
NE1-5M	-	-	-	-	-	-	-	-
NE2-5M	2.5	8.0	1.0	2.5	6.0	4.0	6.0	6.0
NE3-5M	6.5	1.0	2.0	8.0	5.0	3.5	6.5	3.5
TN1-5M	7.0	8.0	2.0	1.0	4.5	6.0	3.0	4.5
TN2-5M	1.0	2.5	2.5	8.0	5.0	7.0	5.0	5.0
SOMME S _D	36.0	50.0	26.5	43.0	59.5	63.5	58.0	59.5
RANG R _D	2.0	4.0	1.0	3.0	6.5	8.0	5.0	6.5
QU1-1H	-	-	-	-	-	-	-	-
QU2-1H	3.0	4.0	1.0	2.0	6.0	8.0	6.0	6.0
QU3-1H	3.0	7.5	2.0	1.0	5.5	7.5	4.0	5.5
QU4-1H	6.0	1.0	7.0	8.0	4.0	2.0	4.0	4.0
QU5-1H	7.0	8.0	6.0	3.0	4.5	4.5	1.0	2.0
QU6-1H	4.0	1.0	2.0	3.0	5.5	7.5	7.5	5.5
NB1-1H	3.0	4.0	2.0	1.0	7.0	8.0	5.0	6.0
NB2-1H	1.0	4.0	3.0	2.0	6.0	6.0	8.0	6.0
NB3-1H	6.0	8.0	7.0	5.0	2.0	2.0	4.0	2.0
NB4-1H	2.0	8.0	3.0	1.0	4.5	4.5	6.5	6.5
NE1-1H	5.0	8.0	7.0	6.0	3.5	3.5	2.0	1.0
NE2-1H	-	-	-	-	-	-	-	-
NE4-1H	2.0	7.0	3.0	1.0	7.0	7.0	4.0	5.0
TN2-1H	4.0	2.0	1.0	3.0	7.0	5.0	7.0	7.0
SOMME S _D	46.0	62.5	44.0	36.0	62.5	75.5	59.0	56.5
RANG R _D	3.0	6.5	2.0	1.0	6.5	8.0	5.0	4.0
QU1-24H	3.0	6.0	2.0	1.0	6.0	6.0	6.0	6.0
QU2-24H	-	-	-	-	-	-	-	-
QU3-24H	1.0	6.5	3.0	2.0	6.5	4.0	6.5	6.5
QU4-24H	3.0	4.0	2.0	1.0	6.5	8.0	5.0	6.5
QU5-24H	3.0	5.0	2.0	1.0	7.0	8.0	4.0	6.0
QU6-24H	1.0	4.0	3.0	2.0	8.0	7.0	5.5	5.5
NB1-24H	2.5	4.0	2.5	1.0	6.5	8.0	5.0	6.5
NB2-24H	2.0	4.0	3.0	1.0	7.0	8.0	5.0	6.0
NB3-24H	3.0	8.0	2.0	1.0	6.0	7.0	4.0	5.0
NB4-24H	1.0	8.0	3.5	2.0	6.5	6.5	3.5	5.0
NE1-24H	2.0	8.0	3.0	1.0	6.0	7.0	4.0	5.0
NE2-24H	6.0	8.0	7.0	5.0	1.0	2.5	4.0	2.5
NE3-24H	5.0	8.0	7.0	6.0	3.0	4.0	2.0	1.0
NE4-24H	2.5	4.0	2.5	1.0	6.5	8.0	5.0	6.5
TN1-24H	1.0	8.0	6.0	2.5	4.0	2.5	7.0	5.0
TN2-24H	-	-	-	-	-	-	-	-
SOMME S _D	36.0	82.5	48.5	27.5	80.5	86.5	66.5	73.0
RANG R _D	2.0	7.0	3.0	1.0	6.0	8.0	4.0	5.0
SOMME S _D	118.0	195.0	119.0	106.5	202.5	215.5	183.5	189.0
RANG R _D	2.0	6.0	3.0	1.0	7.0	8.0	4.0	5.0

- D'après le tableau 7.11, la distribution Pearson type 3 aurait conduit à de bons résultats.

Pour résumer les résultats de l'analyse du CIA, mentionnons que la distribution log-gamma donne les meilleurs résultats suivie des distributions Gumbel et log-normale. De plus, l'amélioration apportée par des modèles de distribution à 3 paramètres n'est pas suffisamment nette pour justifier leur utilisation face à une distribution adéquate à deux paramètres.

7.7 TEST DU RAPPORT DES VRAISEMBLANCES MAXIMALES

Le test du rapport des vraisemblances maximales indique si une distribution est significativement meilleure qu'une autre distribution. Les deux distributions doivent être de la même famille et l'estimation des paramètres doit être faite par la méthode du maximum de vraisemblance. L'équation 3.17 indique le lien entre les valeurs du CIA et la statistique t du test. Soient les hypothèses suivantes:

H_0 : les observations proviennent de la loi F_0 (2 paramètres)

H_1 : les observations proviennent de la loi F_1 (3 paramètres)

L'hypothèse H_0 n'est pas rejetée à un niveau de signification de 5% (équation 3.17) dans le cas où $AIC(F_0) - AIC(F_1) + 2 < 3.84$.

À partir du tableau 7.11, le test du rapport des vraisemblances maximales est appliqué sur les paires de distributions suivantes:

a) H_0 : Log-gamma ($m=0$)

H_1 : Log-Pearson type 3 ($m \neq 0$)

- La distribution Log-gamma est rejetée 4 fois sur 40.

b) H_0 : Log-normale LN2 ($t=0$)

H_1 : Log-normale à 3 paramètres LN3 ($t \neq 0$)

- La distribution log-normale LN2 est rejetée 4 fois sur 43

- c) H_0 : Gumbel ($k=0$)
 H_1 : GEV ($k \neq 0$)
 - La distribution Gumbel est rejetée 2 fois sur 42.

- d) H_0 : Gamma ($s=1$)
 H_1 : Gamma généralisée ($s \neq 1$)
 - La distribution Gamma est rejetée 9 fois sur 43.

- e) H_0 : Gamma ($m=0$)
 H_1 : Pearson type 3 ($m \neq 0$)
 - La distribution gamma est rejetée 5 fois sur 36.

Le test du rapport des vraisemblances maximales montre qu'en générale les distributions à deux paramètres ne sont pas rejetées face à l'hypothèse alternative. La distribution Gumbel est la moins souvent rejetée et la distribution gamma est la plus souvent rejetée.

7.8 CONCLUSION

Différentes distributions et différentes méthodes ont été utilisées pour ajuster des séries annuelles de maxima. Ces échantillons provenant de diverses régions de l'est du Canada ont été examinés afin de vérifier les hypothèses d'indépendance et d'homogénéité et de s'assurer de l'absence de valeurs singulières. Les tests effectués pour comparer les distributions nous conduisent aux conclusions suivantes:

- Parmi les distributions à deux paramètres, les distributions log-normale, log-gamma et Gumbel (méthode des moments) représentent le mieux les séries annuelles de maxima pour les périodes de retour allant de 2 à 50 ans. Une légère préférence est accordée aux distributions log-normale et log-gamma en raison du mauvais comportement de la distribution Gumbel pour les

durées de 5 minutes et 1 heure.

- L'ensemble des distributions à trois paramètres ajuste bien les échantillons. Une faible supériorité est remarquée pour les distributions Pearson type 3 (méthode des moments) et log-normale à 3 paramètres (méthode des moments).
- Les distributions à trois paramètres n'améliorent pas suffisamment l'ajustement pour justifier leur utilisation. De plus, l'emploi d'une loi à plus de 2 paramètres pour des échantillons d'aussi petite taille n'est généralement pas recommandé.
- Parmi les distributions à 3 paramètres, la loi Pearson type 3 présente de l'intérêt en raison de la solution directe apportée par la méthode des moments.
- Il n'y a pas de raison évidente laissant croire que la même distribution ne peut pas être utilisée pour différentes durées. En effet, les résultats sont assez similaires pour toutes les durées.

À plusieurs occasions, les différences rencontrées entre deux méthodes d'estimation pour une même distribution sont plus grandes que les différences rencontrées entre deux différentes distributions. Il s'agit d'une raison importante justifiant la recherche de la meilleure méthode d'estimation des paramètres pour chacune des lois.

La meilleure méthode d'estimation choisie pour chacune des distributions dépend du critère de comparaison. Dans notre cas, l'attention est portée sur la queue droite de la distribution (région des grandes périodes de retour). Un autre critère de comparaison aurait pu conduire à des résultats différents. Il en va de

même pour la sélection des meilleures distributions. Habituellement, la distribution ou la méthode d'estimation la plus adéquate dépend de la période de retour considérée et du coefficient d'asymétrie (forme) de la population.

CHAPITRE 8

CONCLUSION GÉNÉRALE ET DISCUSSION

L'étude statistique des variables de précipitation tient une place importante dans plusieurs domaines pratiques et théoriques. Il existe un grand nombre de distributions statistiques pouvant représenter ces variables. Cette étude a consisté en l'identification de la loi la plus adéquate pour représenter quelques unes d'entre elles.

Dans une première partie, on a présenté les bases théoriques des distributions et de l'ajustement. Par la suite, une revue des distributions retrouvées en pratique a permis de mettre en évidence la diversité des lois employées. Finalement, on a effectué l'ajustement proprement dit des épisodes pluvieux et secs, des hauteurs de précipitation et des valeurs extrêmes de précipitation. Cette dernière phase avait pour but de déterminer la loi la plus adéquate pour chacune de ces variables.

Nous avons sélectionné des stations climatiques du SEA réparties à travers le Québec et les Provinces Atlantiques pour faire une étude comparative des lois. La première étape a consisté à vérifier les hypothèses de base que doivent respecter les échantillons (indépendance, homogénéité et absence de valeur singulière). Nous avons ensuite sélectionné les lois et les méthodes d'estimations des paramètres pouvant représenter les échantillons. Puis nous avons effectué l'ajustement à l'aide du logiciel AJUSTE 1.1 ou avec des programmes en langage GAUSS. La troisième étape a permis de déterminer la distribution la plus adéquate pour représenter les échantillons. Les critères de comparaison que nous avons utilisés sont le critère d'Akaike, le test du rapport des vraisemblances maximales, le test du chi-2 et la comparaison entre les valeurs observées et théoriques.

Le tableau 8.1 présente une synthèse des résultats obtenus.

Tableau 8.1 Synthèse de l'étude de représentation des variables de précipitation par des lois statistiques

Variables	Distributions retenues		Commentaires	
Épisode pluvieux	BNT		La distribution GEO donne également de bons résultats	
Épisode sec	GEO			
Cycle (Tp-Ts)	BNT-GEO			
Hauteur de pluie journalière mensuelle	3 paramètres	2 paramètres	Le modèle GEO-GEO est aussi acceptable	
	EXM	WE		La distribution à 3 paramètres est significativement meilleure que les distributions à 2 paramètres
	LP3, GG	GA, WE		Les distributions à 3 paramètres ne sont pas significativement meilleures que les distributions à 2 paramètres
Séries annuelles de maxima (5 min, 1 et 24 h)	P3	LN, LG, GU	Les distributions à 3 paramètres ne sont pas significativement meilleures que les distributions à 2 paramètres	

Le tableau 8.1 montre bien combien il est difficile de distinguer une seule distribution pour représenter une variable donnée. Habituellement, deux ou trois distributions ajustent bien l'ensemble des échantillons. Le choix final dépend de l'utilisation future de nos résultats. En particulier, l'utilisation d'une distribution à 3 paramètres plutôt qu'une distribution à 2 paramètres augmente la variance du quantile x_T souvent employé dans la conception d'ouvrages hydrauliques.

L'étude des variables de précipitation et climatiques demeure ouverte. Il y reste beaucoup à faire, notamment en ce qui concerne le choix d'une distribution produisant une variance minimale des quantiles avec un ajustement précis. Par exemple, une étude de la variance des quantiles peut être menée à l'aide des méthodes "bootstrap" et "jackknife". Ces méthodes sont utiles dans le cas des petits échantillons, tels que ceux rencontrés dans les séries annuelles de maxima, pour lesquels la variance théorique asymptotique est approximative.

D'un autre côté, la nouvelle version du logiciel AJUSTE (2) va contenir plus de tests pour vérifier les critères statistiques et plus de lois. Spécialement, l'addition des lois de Halphen (type A, B et B⁻¹) devrait offrir de nombreuses possibilités nouvelles dans le domaine de la météorologie.

RÉFÉRENCES

- AKAIKE, H. (1974). A new look at the statistical model identification. *IEEE Trans. Autom. Control*, AC-19 (6), pp. 716-723.
- ALEXANDER, K. (1981). Determination of rainfall duration statistics for rain-out models from daily records. *Wat. Res. Res.*, 17(3), pp. 521-528.
- BAGHIRATHAN, V.R. et E.M. SHAW (1978). Rainfall depth-duration-frequency studies for Sri Lanka. *J. Hydro.*, 37(3), pp. 223-239.
- BARDSLEY, W.E. (1984). Note on the form of distributions of precipitation totals. *J. Hydro.*, 73, pp. 187-191.
- BENSON, M.A. (1968). Uniform flood-frequency estimating methods for federal agencies. *Wat. Res. Res.*, 4(5), pp. 891-908.
- BERNIER, J. et D. FANDEUX (1970). Théorie du renouvellement - Application à l'étude statistique des précipitations mensuelles. *Revue de Statistique Appliquée*, 18(2), pp. 75-87.
- BOBÉE, B. et F. ASHKAR (1988). The generalized method of moments applied to the log-Pearson type 3 distribution. *Journal of Hydrolic Engineering*, ASCE, 114(8), pp 899-909.
- BOBÉE, B. et F. ASHKAR (1991). *The Gamma Family and Derived Distributions Applied in Hydrology*. Water Resources Publications, Littleton, Colorado, 203 p.
- BOBÉE, B. et R. ROBITAILLE (1977). The use of the Pearson type 3 and Log Pearson type 3 distributions revisited. *Wat. Res. Res.*, 13(2), pp. 427-443.
- BONSER, J.D., T.E. UNNY et K. SINGHAL (1985). A marked Poisson process of summer rainfall in southern Ontario. *Can. J. Civ. Eng.*, pp. 886-898.
- BRASS, W. (1958). Simplified methods of fitting the truncated negative binomial distribution. *Biometrika*, 45, pp. 59-68.
- BUISHAND, T.A. (1978). Some remarks on the use of daily rainfall models. *J. Hydrol.*, 36, pp. 295-308.
- BUISHAND, T.A. et G.R. DEMARÉ (1990). Estimation of the annual maximum distribution from samples of maxima in separate seasons. *Stochastic Hydrol.*

Hydraul., 4, pp. 89-103.

CAO, C. (1974). A contribution to statistical depth-duration-frequency analysis. *J. Hydrol.*, 22(1/2), pp. 109-129.

CHIN, E.H. (1977). Modelling daily precipitation occurrence process with Markov chain. *Wat. Res. Res.*, 13(6), pp. 949-956.

CHOW, V.T. (1962). *Handbook of Applied Hydrology*. McGraw-Hill, New York.

COLLIN, E. et J.A BEDEL (1980). Essais d'ajustements régionaux sur les intensités de pluie. *La météorologie*, VI^e série, No. 20-21, pp. 151-160.

CTGREF (1978). *Crue et assainissement - Analyse des pluies de 1 à 10 jours sur 300 postes métropolitains*. Division Hydrologie-Hydraulique Fluviale, Antony.

CUNNANE, C. (1978). Unbiased plotting positions - A review *J. Hydrol.*, 37, pp 205-222.

DICKINSON, T. (1977). Rainfall intensity-frequency relationships from monthly extremes. *J. Hydrol.*, 35(1/2), pp 137-145.

DUBANT, C., MAIRE G. et J. SCHWARTZ (1980). Analyse des fortes pluies de 1H à 48H pour une quinzaine de postes alsaciens. *La météorologie*, VI^e série, No. 20-21, pp. 127-137.

EAGLESON, P.S. (1978) Climate, soil, and vegetation, 2, The distribution of annual precipitation derived from observed storm sequences *Wat. Res. Res.*, 14(5), pp. 713-721.

EAGLESON, P.S. (1981). Some limiting forms of the Poisson of annual station precipitation. *Wat. Res. Res.*, 17(3), pp. 752-757.

EGBUNIWE, N. (1975). Generation of hourly rainfall. *Wat. Res. Bull.*, 11(4), pp. 706-713.

EVERITT, B.S. et D.J. HAND (1981). *Finite Mixture Distributions*. Chapman and Hall, London.

FISHER, R.A. et TIPPET, L.H.C. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Proceedings of the Cambridge Philosophical Society*, 24, pp. 180-190.

FITZGERALD, D.L. (1989). Single station and regional analysis of daily rainfall extremes. *Stochastic Hydrol. Hydraul.*, 3, pp. 281-292.

- GABRIEL, K.R., et J. NEUMANN (1962). A Markov chain model for daily rainfall occurrences at Tel-Aviv. *Quart. J. Roy. Meteorol. Soc.*, 88, pp. 90-85.
- GIBRAT, R. (1931). *Les inégalités économiques*, Paris, Libraire du Recueil, Sirey.
- GRAYMAN, W.M. et P.S. EAGLESON (1969). Streamflow record length for modelling catchment dynamics. Hydrodinam. Lab., *Rep. 114*, Dep. of Civil Eng., Mass. Inst. of Technol., Cambridge, Mass.
- GREENWOOD, J.A., LANDWERH, J.M., MATALAS, N.C. et J.R. WALLIS (1979). Probability weighted moments: definition and relation to parameters of several distributions expressible in inverse form. *Wat. Res. Res.*, 15, pp. 1049-1054.
- GRUBBS, F. et G. BECK (1972). Extension of sample sizes and percentage points for significance tests of outlying observations. *Technometrics*, 14(4), pp 847-854.
- HERSHFIELD, D.M (1970). The frequency and intensity of wet and dry seasons and their interrelationships. *Wat. Res. Bull.*, 6(1), pp. 87-93.
- HILL, B.M (1963). The three-parameter log normal distribution and Bayesian analysis of a point-source epidemic. *J. Am. Stat. Assoc.*, 58, pp. 72-84.
- HOGG, W.D. et D.A CARR (1985). *Atlas de la fréquence des pluies au Canada*. Publication du Programme climatologique canadien, Service de l'environnement atmosphérique, Downsview, Ont.
- HUFF, F.A. et J.C. NEILL (1959). Comparison of several methods for rainfall frequency analysis. *J. Geophys. Res.*, 64(5), pp. 541-547.
- INRS-EAU (1992). Ajuste 1.1 - Manuel D'utilisation. *Rapport interne*, 67 p.
- JENKINSON, A.F. (1955). The frequency distribution of the annual maximum (or minimum) of meteorological elements. *Quart. J. Roy. Meteor. Soc.*, 81, pp. 158-171.
- JOHNSON, L.J. et S. KOTZ (1969). *Discrete Distributions*, Wiley, New York, 328 p.
- KAPPUS, U., J.M. BLEEK et S.H. BLAIR (1978). Rainfall frequencies for the Persian Gulf coast of Iran. *Hydro. Sci. Bull.*, 23(4), pp. 85-90.
- KRUSKAL, W.H. et W.A. WALLIS (1952). Use of ranks in one-criterion variance analysis. *J. Am. Stat. Assoc.*, 47, pp. 583-621.
- KULLBACK, S. (1959). *Information and Statistics*, Wiley, New York.

- KULLBACK, S. et R.A. Leibler (1951). On information and sufficiency, *Ann. Math. Stat.*, 22, pp. 79-86.
- LAHAYE J.P. (1980). Étude des hauteurs de pluie exceptionnelles en un et plusieurs jours consécutifs en Haute-Volta. *La météorologie*, VI^e série, No. 20-21, pp. 191-197.
- LAVABRE, J. (1980). La pluviométrie du bassin expérimental du Real Collobrier. *La météorologie*, VI^e série, No. 20-21, pp. 119-126.
- LAWLEY, D.N. (1956). A general method for approximating to the distribution of likelihood ratio criteria. *Biometrika*, 43, pp. 295-303.
- LINSLEY, R.K., KOHLER, M.A. et PAULHUS, J.L.H (1975). *Hydrology for Engineers*. McGraw-Hill, New York.
- LOBERT, A. (1967). Modèle probabiliste de base pour les pluies dans le bassin de l'Allier. *E.D.A*, 45.
- MAJUMBAR, K.C. et R.P. SAWHNEY (1965). Estimates of extreme values by different distribution functions. *Wat. Res. Res.*, 1(3), pp. 429-434.
- MANN, H.B. et D.R. WHITNEY (1947). On the test of whether one of two random variables is stochastically larger than the other. *Ann. Math. Statist.*, 18, pp. 50-60.
- MARKOVIC, R.D. (1965). Probability functions of best fit to distributions of annual precipitation and runoff. *Hydrology Papers*, #8, Colorado State University, Fort Collins, Colo., 33 p.
- MASSON, J.M. (1980). Analyse d'une longue série pluviographique Montpellier Bel Air 1920-1971. *La météorologie*, VI^e série, No. 20-21, pp. 105-117.
- MIELKE, P.W. (1973). Another family of distributions for describing and analysing precipitation data. *J. Appl. Meteorol.*, 15(2), pp. 275-280.
- MILLER, J.F., R.H. FREDERICK et R.J. TRACEY (1973). *Precipitation-frequency Atlas of the western United States*. Vol.I-Vol.XI, NOAA Atlas 2, National Weather Service, US Dept. of Commerce, Silver Spring. Md.
- NRCC (1989). *Hydrology of Floods in Canada: A Guide to Planning and Design*. National Research Council of Canada, Ottawa, 245 p.
- NGUYEN, V.T.V (1984a). A stochastic description of temporal daily rainfall patterns. *Canadian J. of Civil Eng.*, 11(2), pp. 234-238.

- NGUYEN, V.T.V. (1984b). On stochastic characterization of temporal storm patterns
Wat. Sci. Tech., 16, pp. 147-153.
- NGUYEN, V.T.V. (1987). On the study of a probability distribution for hourly rainfall amounts. *Fourth International Conference on Urban Storm Drainage*, Lausanne, Suisse.
- NGUYEN, V.T.V. et A. MAYABI (1990). Probabilistic analysis of summer daily rainfall for the Montreal region. *Canadian Wat. Res. J.*, 15(3), pp. 65-80
- NGUYEN, V.T.V. et J. ROUSSELLE (1981). A stochastic model for the time distribution of hourly rainfall depth. *Wat. Res. Res.*, 17(2), pp. 399-409
- OZTURK, A. (1981). On the study of a probability distribution for precipitation totals. *J. Appl. Meteor.*, 20, pp. 1499-1505.
- PERREAULT, L. et B. BOBÉE (1992a). Loi normale. INRS-Eau, *Rapport Scientifique no. 352*.
- PERREAULT, L. et B. BOBÉE (1992b). Loi généralisée des valeurs extrêmes. INRS-Eau, *Rapport Scientifique no. 350*.
- PERREAULT, L. et B. BOBÉE (1992c). Loi Weibull à deux paramètres. INRS-Eau, *Rapport Scientifique no. 351*.
- PHIEN, H.N. et T. JIVAJIRAJAH (1984). Fitting annual rainfall and annual streamflow by two transformed gamma distributions. *Water S.A.*, 10(2), pp. 65-74.
- QUÉLENNEC, R.E. (1970). Étude du régime pluviométrique des bassins de la Charente et de la Seudre. *Agence de bassin Adour-Garonne*, 92 pages.
- QUÉLENNEC, R.E. (1973). Contribution à l'étude probabiliste des phénomènes pluvieux, Application aux bassins de la Charente et de la Seudre. *La houille blanche*, No. 1, pp. 21-33.
- RAO, D.V. (1980). Log-Pearson type 3 distribution: method of mixed moments. *Journal of the Hydraulics Division, ASCE*, 106(6), pp. 999-1019.
- RAO, A.R. et B.T. CHENCHAYYA (1974). Probabilistic analysis and simulation of the short time increment rainfall process. *Wat. Res. Res. Center, Rept. 55*, Purdue Univ., West Lafayette, Indiana.
- RAO, S.G. et A.R. RAO (1986). Run analysis of rainfall data affected by urbanization. *Nord. Hydro.*, 17, pp. 47-64.

- RAUDKIWI, A.J. et N. LAWGUN (1970). Synthesis of urbain rainfall. *Wat. Res. Res.*, 6(2), pp. 455-464.
- REICH, B.M. (1963). Short-duration rainfall-intensity estimates and other design aids for region of sparce data. *J. Hydro.*, 1(1), pp. 3-28.
- REVIEM, K.J.A. (1983). On the analysis of extreme rainfalls. *J. Hydro.*, 62, pp. 107-117.
- REVIEM, K.J.A. (1984). An initial model of the relationship between rainfall events and daily rainfall. *J. Hydro.*, 75, pp. 357-364.
- RICE, J.A. (1987). *Mathematical Statistics and Data Analysis*. Wadsworth & Brooks/Cole Advanced Books & Software, Pacific Grove, California, 595 p.
- RICHARDSON, C.W. (1982). A comparison of three distributions for the generation of daily rainfall amounts. *Statistical analysis of rainfall and runoff*, Edité par V.P. Singh, Wat. Res. Pub., Littleton, Colo., pp. 67-78.
- ROLDAN, J. et D.A. WOOLHISER (1982). Stochastic daily precipitation models, 1, A comparison of occurrence processes. *Wat. Res. Res.*, 18(5), pp. 1451-1459.
- SAMUELSON, B. (1972). Statistical interpretation of hydrometeorological extreme values. *Nord. Hydrol.*, 3(4), pp. 199-213.
- SEVRUK, B. et H. GEIGER (1981). Selection of distribution types for extremes of precipitation. *World. Meteorological Organization Report 15*.
- SHARMA, T.C. (1987). Modelling extreme rain sequences in Zambia. *J. Hydro.*, 93, pp. 101-111.
- SMITH R.E. et H.A. SHREIDER (1973). Point process of seasonal thunderstorm rainfall, 1, Distribution of rainfall events. *Wat. Res. Res.*, 9(4), pp. 871-884.
- SMITH, R.E. et H.A. SHREIDER (1974). Point processes of seasonal thunderstorms rainfall, 2, Rainfall depth probabilities. *Wat. Res. Res.*, 10(3), pp. 418-423.
- SPIELGE, R.M. (1989). *Formules et tables de mathématiques*. Série Schaum, McGraw-Hill, New York, 272 p.
- STIDD, C.K. (1952). Cube root normal precipitation distribution. *Eos Trans. AGU*, 34(1), pp. 31-35.

- STOL, P.T. (1971). On the decomposition of the extreme value distribution of daily rainfall depths and the derivation of probabilities of compound events. *J. Hydrol.*, 14(3/4), pp. 181-196.
- SUZUKI, E. (1964). Hyper-gamma distribution and its fitting to rainfall data. *Pap. Meteor. Geophys.*, 15, pp. 31-51.
- SUZUKI, E. (1980). A summarized review of theoretical distributions fitted to climatic factors and Markov chain models of weather sequences, with some examples. *Statistical Climatology, Developments in Atmospheric Science*, 13, Elsevier, pp. 1-20.
- THOM, H.C.S. (1959). A time interval distribution for excessive rainfall. *J. Hydraulics Division*, 7, pp. 83-91.
- U.S. WEATHER BUREAU (1961). Rainfall intensity atlas of the United States. *Technical Paper No. 40*, Washington, D.C.
- VAN MONFORT, M.A.J. et J.V. WITTER (1986). The generalized Pareto distribution applied to rainfall depths. *Hydrological Sciences Journal*, 31(2), pp. 151-161.
- WALD, A. et J. WOLFOWITZ (1943). An exact test for randomness in the non-parametric case based on serial correlation. *Ann. Math. Statist.*, 14, pp. 378-388.
- WAYMIRE, E. et V.K. GUPTA (1981). The mathematical structure of rainfall representations, 1, A review of the stochastic rainfall models. *Wat. Res. Res.*, 17(5), pp. 1261-1272.
- WOOLHISER, D.A. et J. ROLDAN (1982). Stochastic daily precipitation models, 2, A comparison of distributions of amounts. *Wat. Res. Res.*, 18(5), pp. 1461-1468.
- WOOLHISER, D.A. et G.G.S. PEGRAM (1979). Maximum likelihood estimation of Fourier coefficients to describe seasonal variations of parameters in stochastic daily precipitation models. *J. Appl. Meteorol.*, 18, pp. 34-42.
- ZUCCHINI, W. et P.T. ADAMSON (1989). Bootstrap confidence intervals for design storms from exceedance series. *Hydrological Sciences Journal*, 34(1), pp. 41-48.

ANNEXE A

ORIGINE DU CRITÈRE D'INFORMATION D'AKAÏKE

Dans cette section, nous décrivons brièvement l'origine et les bases théoriques du critère d'information d'Akaïke. Pour plus de détails, nous vous invitons à consulter Akaïke (1974).

Une approche naturelle pour identifier le meilleur modèle pour représenter un échantillon est de choisir une fonction de risque appropriée $R(\theta, \hat{\theta})$ et de choisir la densité qui minimise le risque. Akaïke (1974) considère l'espérance de la mesure de divergence de Kullback-Leiber (1951) comme fonction de risque:

$$E[W(\theta, \hat{\theta})] \tag{A.1}$$

$$\text{où } W(\theta, \hat{\theta}) = -2 \int f(x; \theta) \ln \left(\frac{f(x; \hat{\theta})}{f(x; \theta)} \right) dx \tag{A.2}$$

Considérons une famille de fonctions de densité $f(x; \theta_p)$ avec $\theta_p \in \mathbb{R}^p$ et la vraie fonction de densité $f(x; \theta_M)$ avec $\theta_M \in \mathbb{R}^M$ ($M \geq p$). De plus, considérons $\hat{\theta}_p$ l'estimateur du maximum de vraisemblance de θ_p . On peut montrer que:

$$W[\theta_M, \hat{\theta}_p] = \|\tilde{\theta}_p - \theta_M\|_J^2 + \|\hat{\theta}_p - \theta_p\|_J^2 \tag{A.3}$$

avec $\|\theta\|_J^2 = \theta' J \theta$ où J est la matrice d'information de Fisher (Kullback, 1959). Ici, $\tilde{\theta}_p \in \mathbb{R}^M$ est la projection de $\theta_p \in \mathbb{R}^p$ dans \mathbb{R}^M .

Pour n grand, la quantité

$$n \|\hat{\theta}_p - \theta_p\|_J^2 \tag{A.4}$$

suit approximativement une distribution chi-2 à p degrés de liberté. D'un autre côté, le rapport des vraisemblances:

$$-2 \sum_{i=1}^n \ln [f(x_i; \hat{\theta}_p) / f(x_i; \hat{\theta}_M)] \tag{A.5}$$

suit approximativement une distribution chi-2 non-centrée à $M-p$ degrés de liberté avec un paramètre de non-centralité de $n \|\tilde{\theta}_p - \theta_M\|_J^2$. À partir des équations (A.4) et (A.5), Akaïke (1974) montre que l'estimateur non-biaisé de l'espérance mathématique

de l'équation (A.1) est:

$$E[W(\theta_M, \theta_p)] = \frac{1}{n} \left[-2 \sum_{i=1}^n \ln[f(x_i; \hat{\theta}_p)] + 2 \sum_{i=1}^n \ln[f(x_i; \hat{\theta}_M)] + 2p - M \right] \quad (A.6)$$

En adoptant l'équation (A.6) comme fonction de risque, elle peut être minimisée en fonction de p afin de trouver la meilleure $f(x; \theta_p)$. Avec M et n fixe, le minimum de l'équation (A.6) ne dépend que du CIA:

$$CIA = -2 \sum_{i=1}^n \ln[f(x_i; \hat{\theta}_p)] + 2p = -2 \ln L + 2p \quad (A.7)$$

ANNEXE B

LA DISTRIBUTION SÉRIE LOGARITHMIQUE:
CAS LIMITE DE LA LOI BINOMIALE NÉGATIVE TRONQUÉE

L'objectif de cette annexe est de montrer que la distribution série logarithmique (équation (2.21a)) est obtenue en prenant la limite, lorsque $r \rightarrow 0$, de la distribution binomiale négative tronquée (équation (2.26a)):

$$\begin{aligned}
 \lim_{r \rightarrow 0} f(x) &= \lim_{r \rightarrow 0} \binom{x+r-1}{x} \frac{b^r (1-b)^x}{1-b^r} \\
 &= \lim_{r \rightarrow 0} \frac{\Gamma(x+r)r}{x! \Gamma(r+1)} \frac{b^r (1-b)^x}{1-b^r} \quad (\text{B.1}) \\
 &= \frac{(1-b)^x}{x!} \lim_{r \rightarrow 0} \frac{\Gamma(x+r)rb^r}{\Gamma(r+1)(1-b^r)}
 \end{aligned}$$

Mais, $\lim_{r \rightarrow 0} \frac{\Gamma(x+r)rb^r}{\Gamma(r+1)(1-b^r)} = \frac{0}{0}$ est indéterminée. Pour lever l'indétermination on applique la règle de l'Hospital:

$$\begin{aligned}
 \lim_{r \rightarrow 0} \frac{\Gamma(x+r)rb^r}{\Gamma(r+1)(1-b^r)} &= \lim_{r \rightarrow 0} \frac{d[\Gamma(x+r)rb^r]/dr}{d[\Gamma(r+1)(1-b^r)]/dr} \\
 &= \lim_{r \rightarrow 0} \frac{\Psi(x+r)rb^r + \Gamma(x+r)(b^r + rb^r \ln(b))}{\Psi(r+1)(1-b^r) - \Gamma(r+1)b^r \ln(b)} \quad (\text{B.2}) \\
 &= \frac{\Gamma(x)}{-\ln(b)} = \frac{(x-1)!}{-\ln(b)}
 \end{aligned}$$

Où $\Psi(\cdot) = \Gamma'(\cdot)$. On déduit ensuite de (B.1) et (B.2) le résultat:

$$\lim_{r \rightarrow 0} f(x) = \frac{(x-1)!(1-b)^x}{-x! \ln(b)} = \frac{1}{-\ln(b)} \frac{(1-b)^x}{x}$$

Ce qui est équivalent à l'équation (2.21) avec $\theta = 1-b$.

ANNEXE C

CHANGEMENT D'UNITÉ DE MESURE AVEC LA LOI LOG-GAMMA

Le but de cette annexe est de démontrer que si la loi log-gamma est applicable à un échantillon exprimé dans un certain système d'unités de mesure elle ne sera pas applicable si l'unité de mesure est changée (par exemple, de mètres en pieds).

Soit l'échantillon x_1, x_2, \dots, x_n exprimé dans le système de mesure (A) et x'_1, x'_2, \dots, x'_n le même échantillon exprimé dans un différent système de mesure (B). Les deux variables aléatoires X et X' sont reliées une à l'autre par:

$$x' = \phi x \quad (C.1)$$

Supposons que la variable X suit une distribution $LG(\alpha, \lambda)$ dont la f.d.p. est donnée par (équation 2.73 avec $m=0$):

$$f_x(x) = \frac{k|\alpha|}{x\Gamma(\lambda)} \exp[-\alpha \log_a(x)] [\alpha \log_a(x)]^{\lambda-1} \quad \begin{array}{l} 1 \leq x < \infty \text{ si } \alpha > 0 \\ 0 \leq x \leq 1 \text{ si } \alpha < 0 \end{array} \quad (C.2)$$

Effectuons maintenant le changement de variable $x' = u(x) = \phi x$. La f.d.p. de la variable aléatoire X' s'écrit:

$$f_{x'}(x') = f_x[u^{-1}(x')] |du^{-1}(x')/dx'| \quad (C.3)$$

où $u^{-1}(x') = x'/\phi$ et $|du^{-1}(x')/dx'| = 1/\phi$. En substituant ces deux derniers termes dans l'équation (C.3), on obtient:

$$\begin{aligned} f_{x'}(x') &= \frac{k|\alpha|}{(x'/\phi)\Gamma(\lambda)} \exp[-\alpha \log_a(x'/\phi)] [\alpha \log_a(x'/\phi)]^{\lambda-1} \frac{1}{\phi} \\ &= \frac{k|\alpha|}{x'\Gamma(\lambda)} \exp[-\alpha(\log_a(x') - m)] [\alpha(\log_a(x') - m)]^{\lambda-1} \quad \begin{array}{l} \phi \leq x < \infty \text{ si } \alpha > 0 \\ 0 \leq x \leq \phi \text{ si } \alpha < 0 \end{array} \end{aligned} \quad (C.4)$$

où $m = \log_a(\phi)$. Donc, la variable X' ne suit pas une distribution $LG(\alpha, \lambda)$ mais une distribution $LP3(\alpha, \lambda, \log_a(\phi))$.

ANNEXE D

EXEMPLES D'AJUSTEMENT SUR PAPIER DE PROBABILITÉ

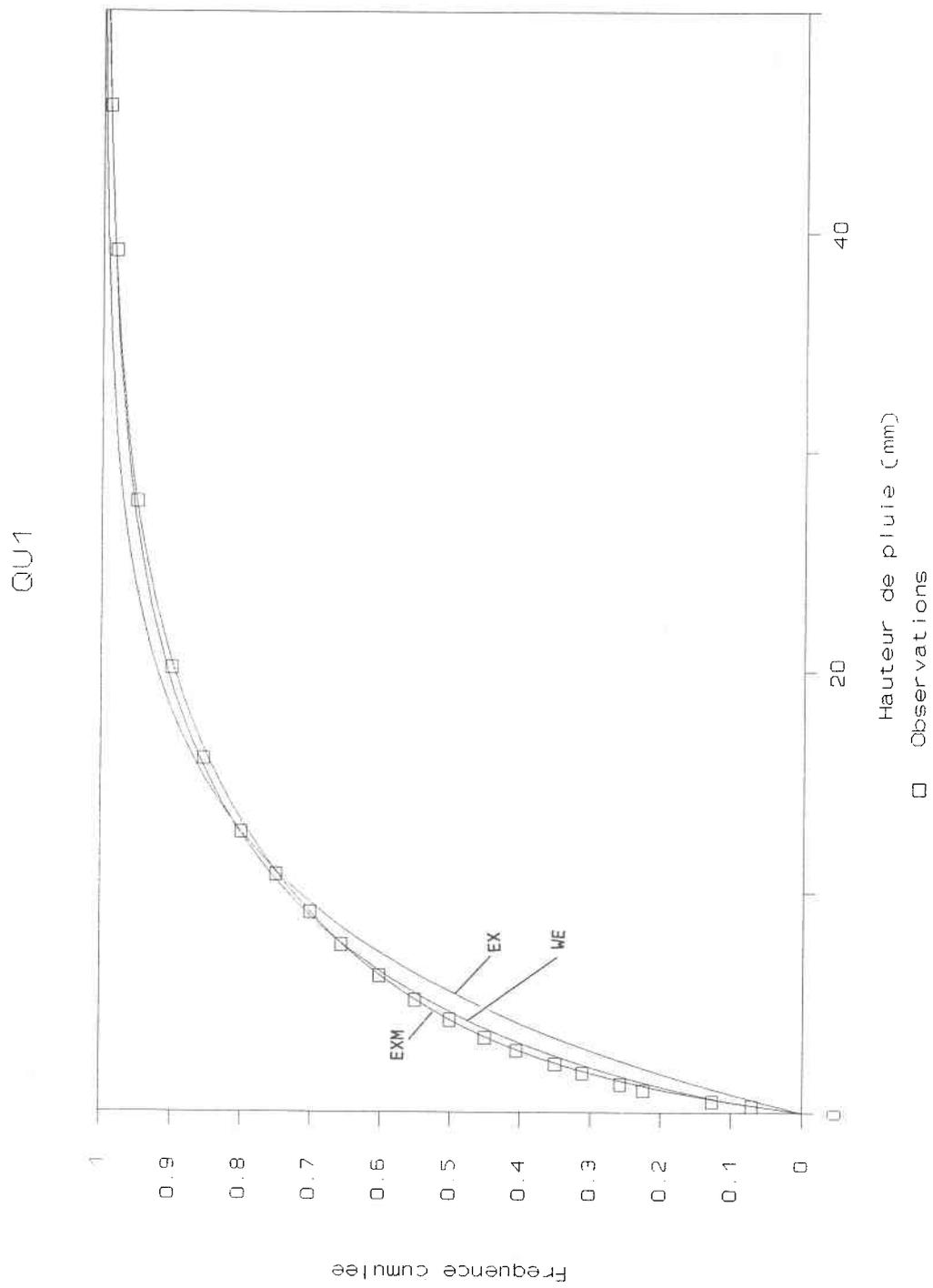


Figure D.1 Ajustement sur papier de probabilité des hauteurs de pluie journalière - station QU1 (EX,WE,EXM)

QU3

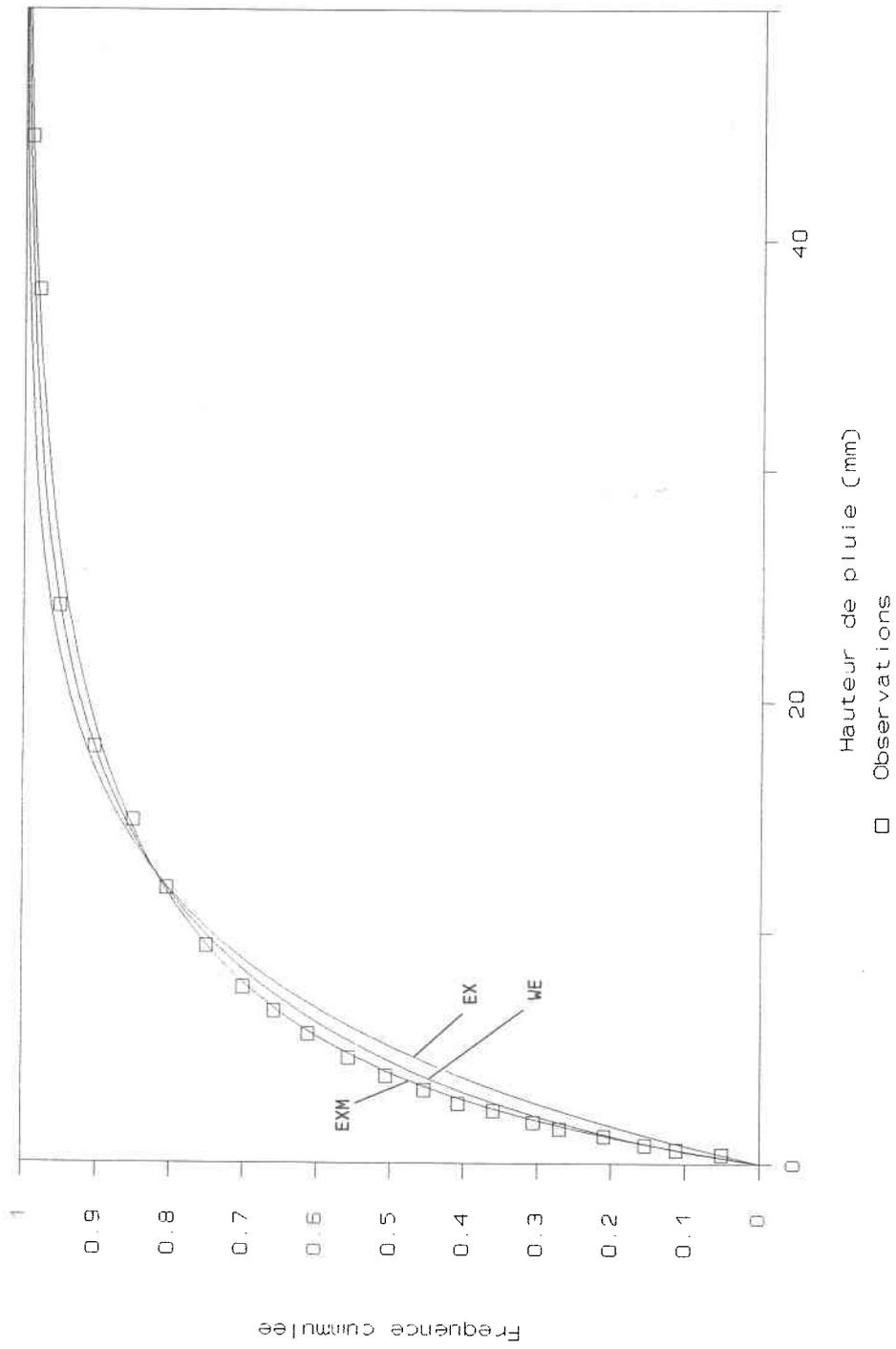


Figure D.2 Ajustement sur papier de probabilité des hauteurs de pluie journalière - station QU3 (EX, WE, EXM)

QU4

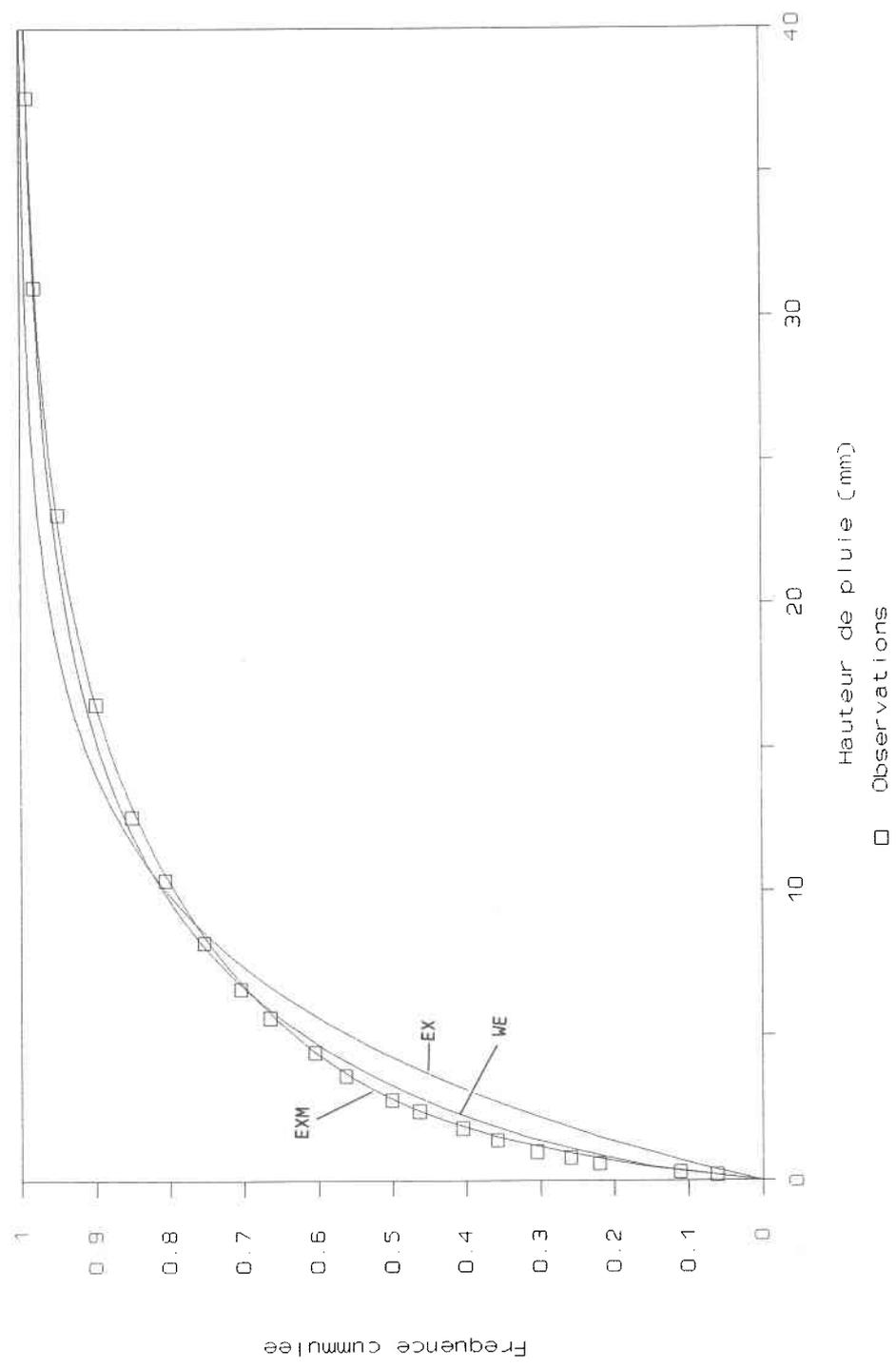


Figure D.3 Ajustement sur papier de probabilité des hauteurs de pluie journalière - station QU4 (EX, WE, EXM)

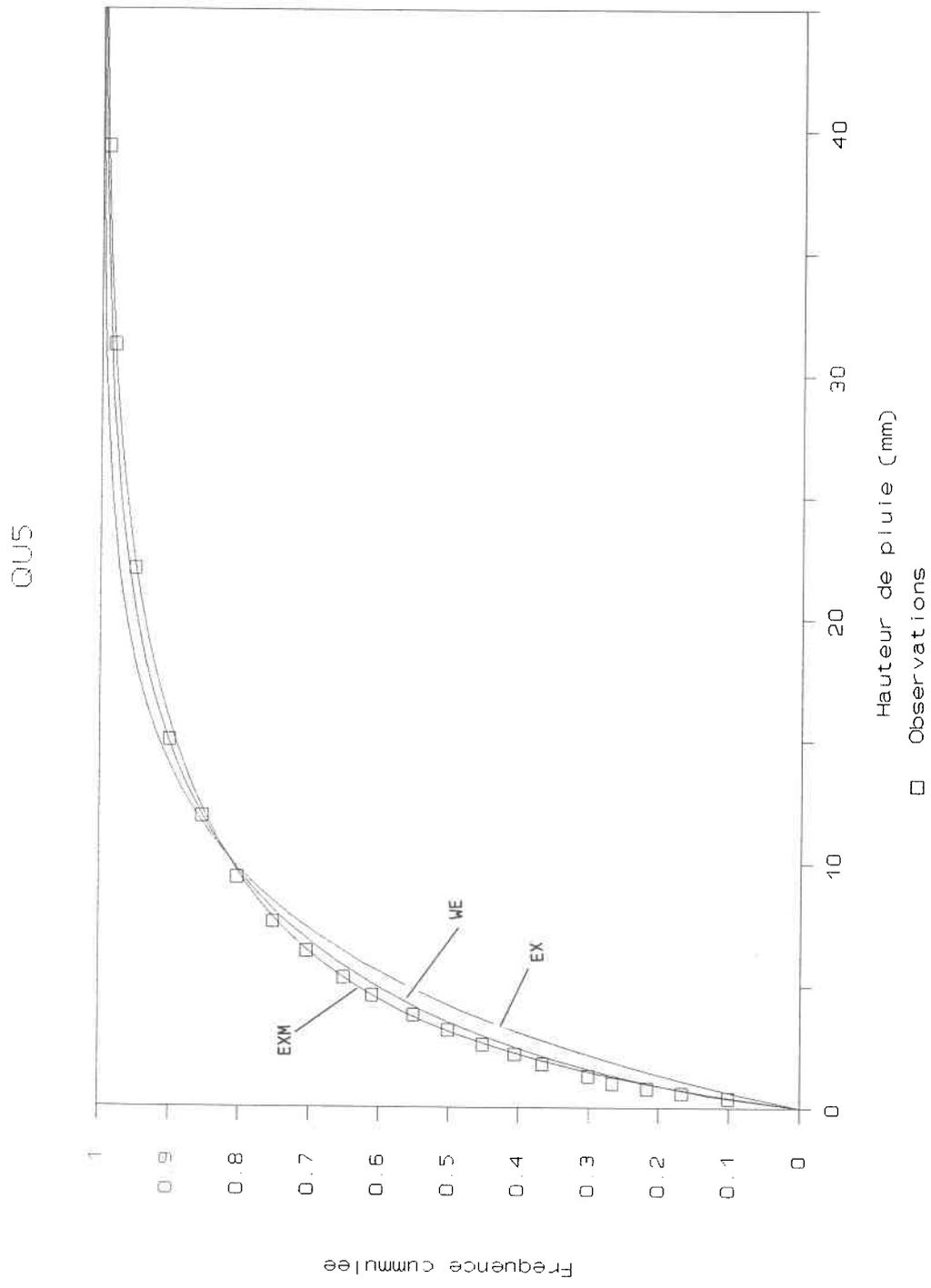


Figure D.4 Ajustement sur papier de probabilité des hauteurs de pluie journalière - station QU5 (EX, WE, EXM)

NB2

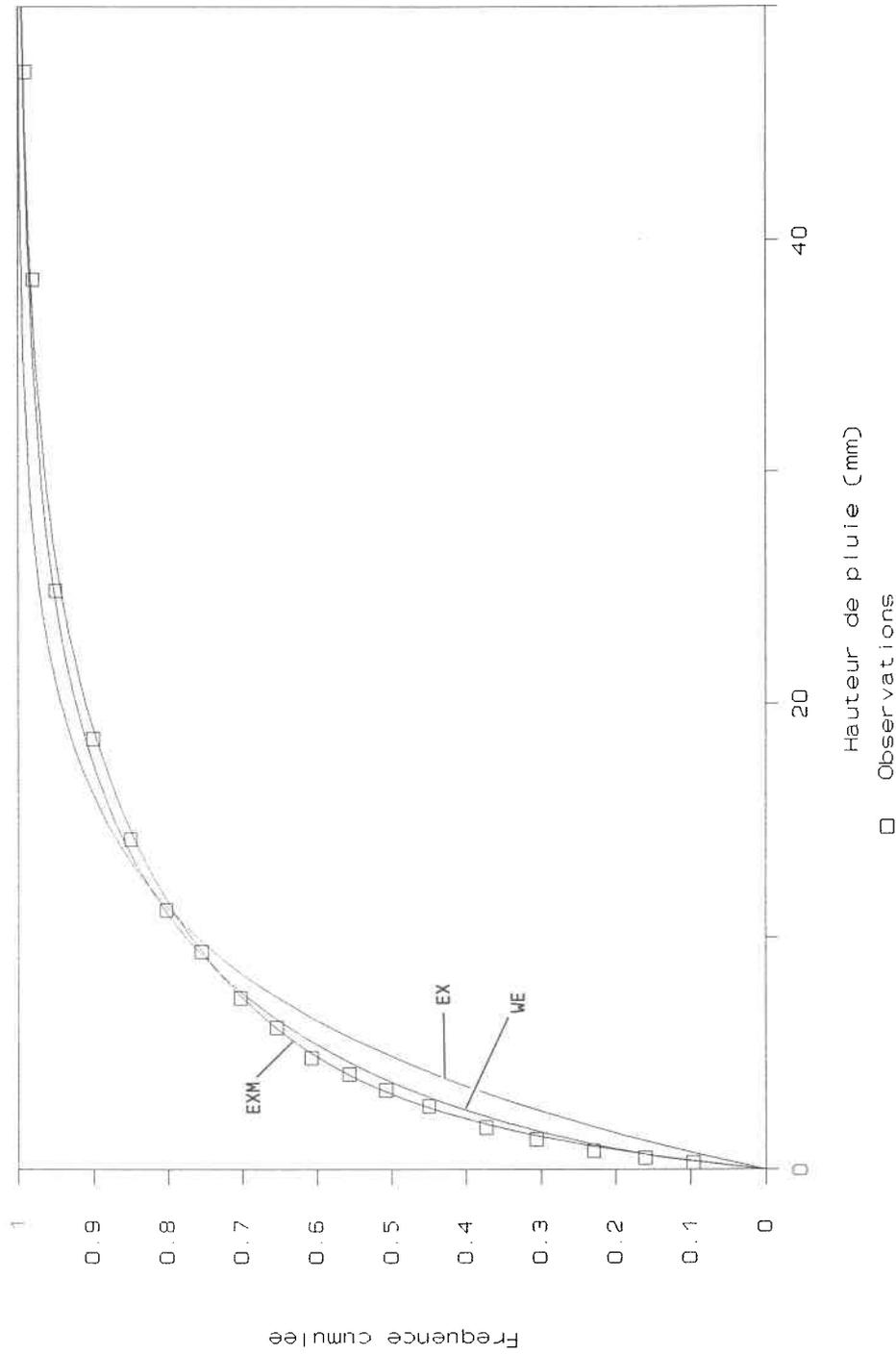


Figure D.5 Ajustement sur papier de probabilité des hauteurs de pluie journalière - station NB2 (EX, WE, EXM)

NB3

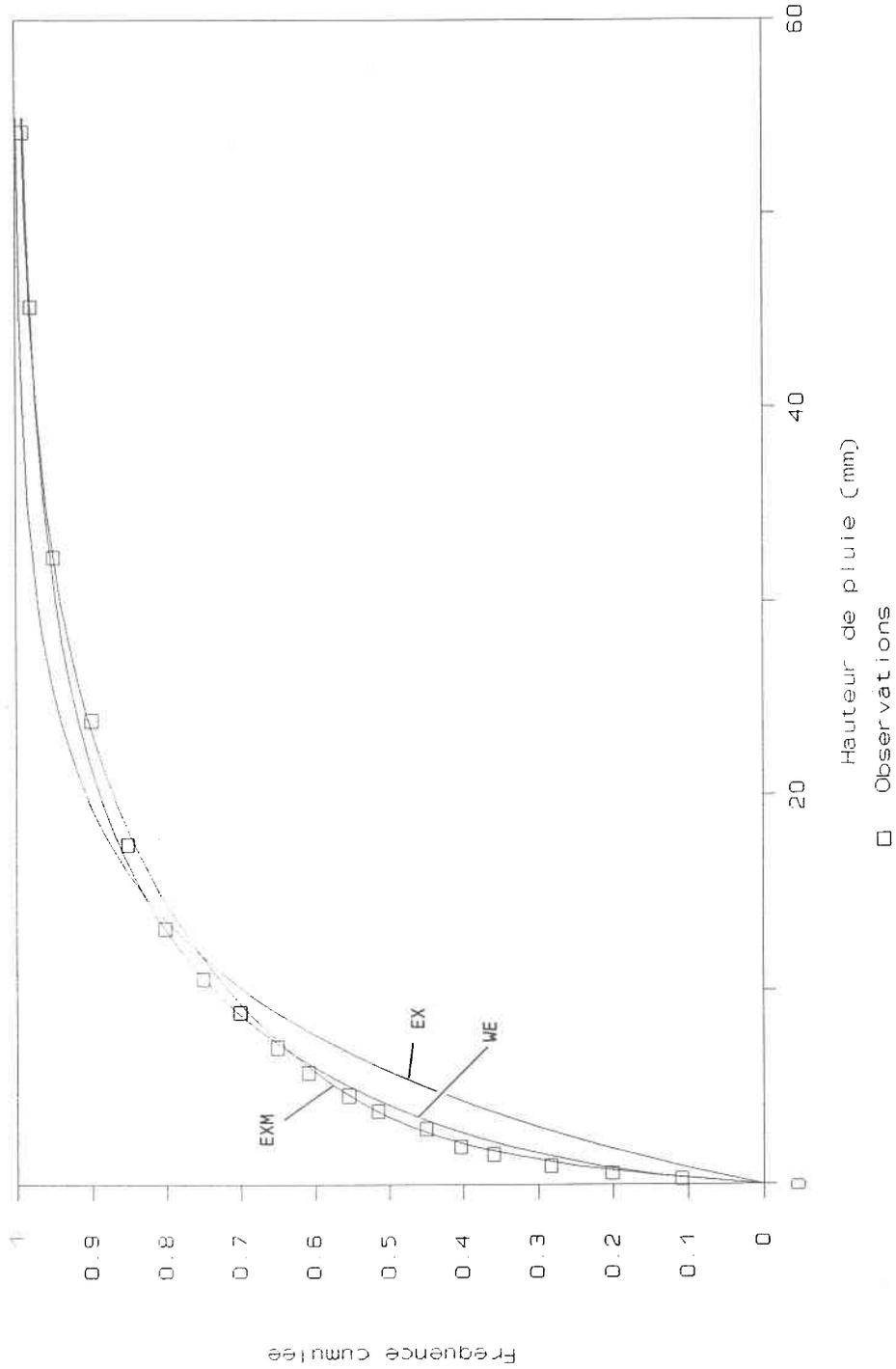


Figure D.6 Ajustement sur papier de probabilité des hauteurs de pluie journalière - station NB3 (EX, WE, EXM)

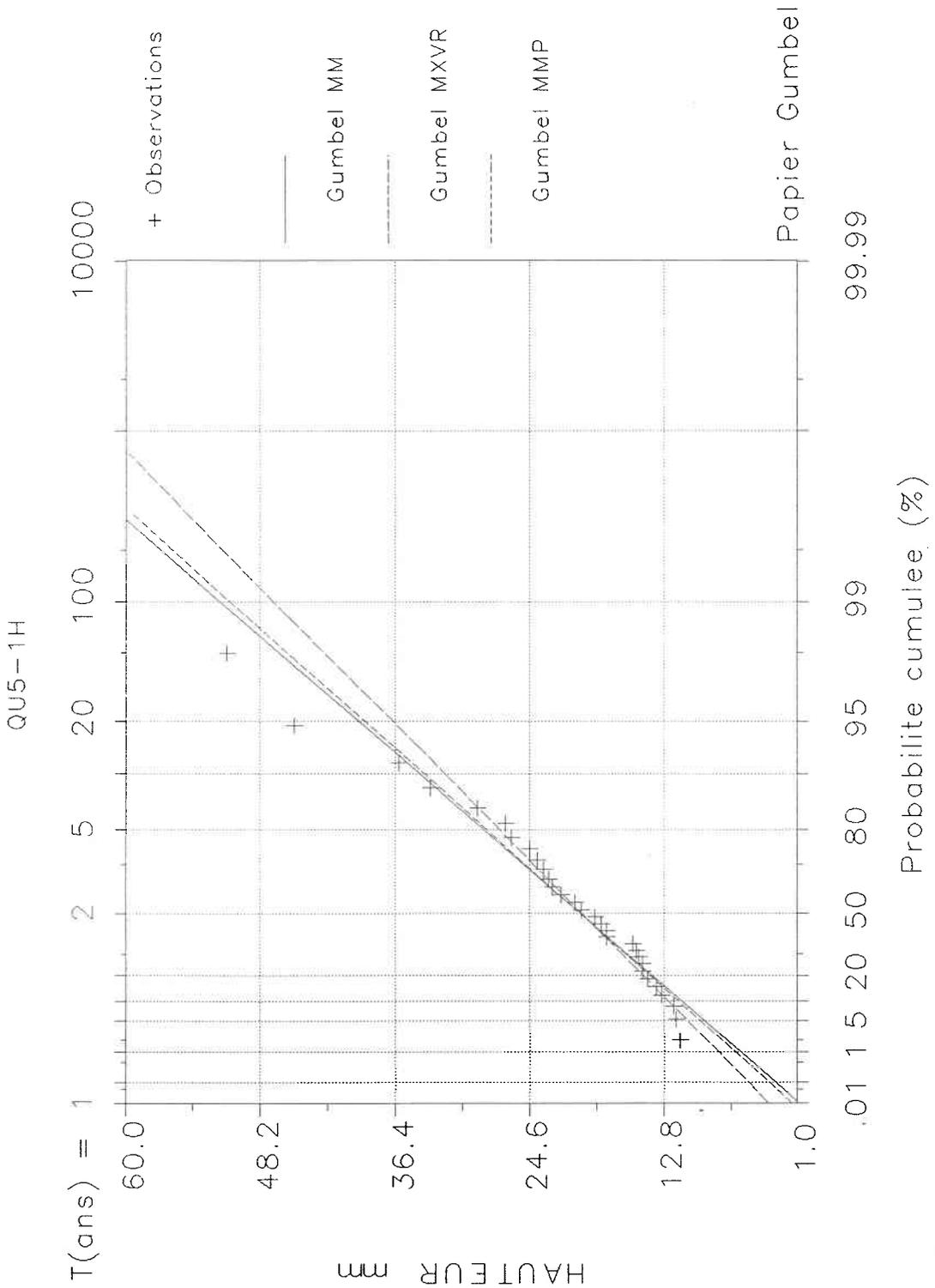


Figure D.7 Ajustement sur papier de probabilité des séries annuelles de maxima - station QU5-1H (GU)

NE4-1H

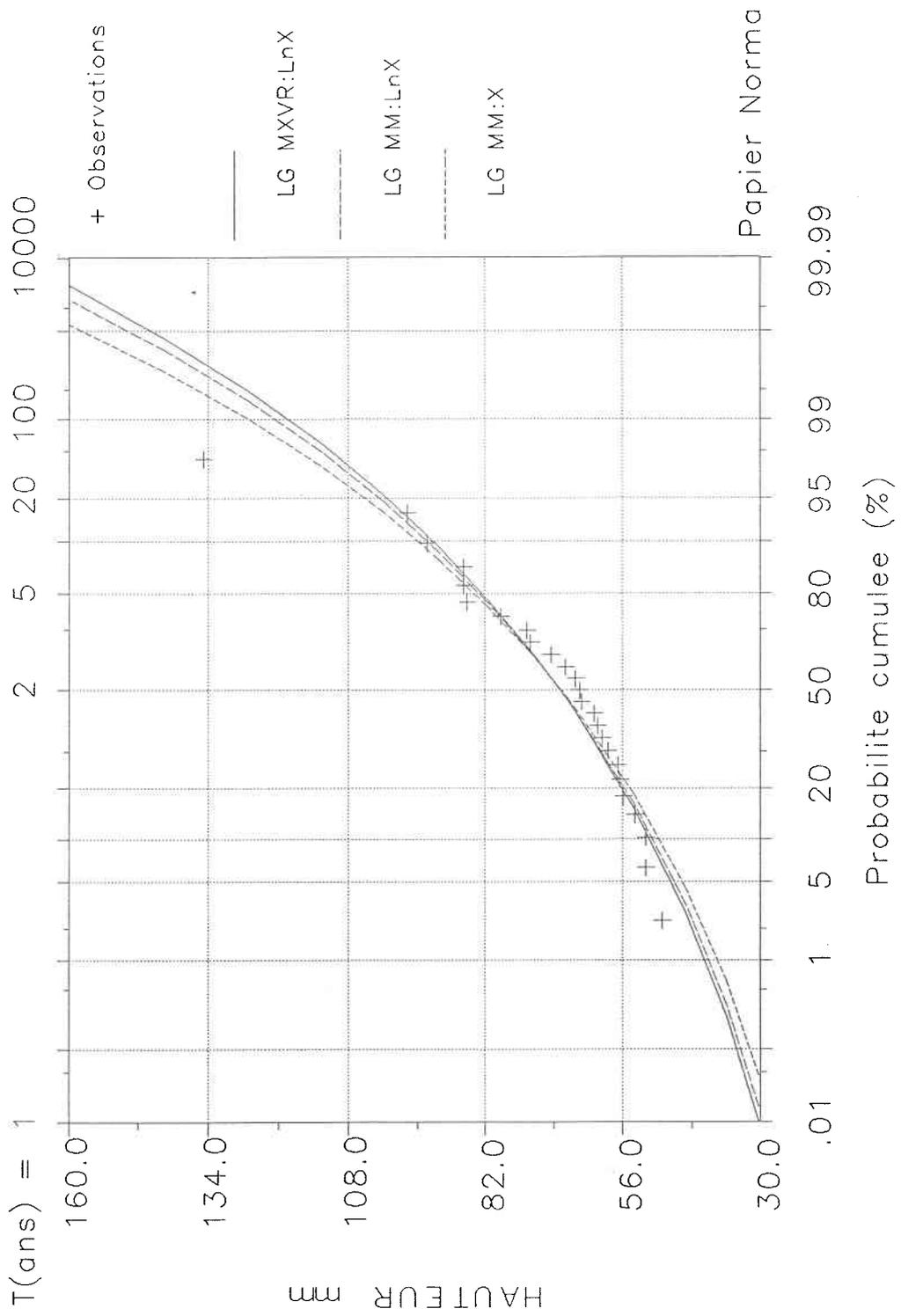


Figure D.8 Ajustement sur papier de probabilité des séries annuelles de maxima - station NE4-24H (IG)

NB3-5M

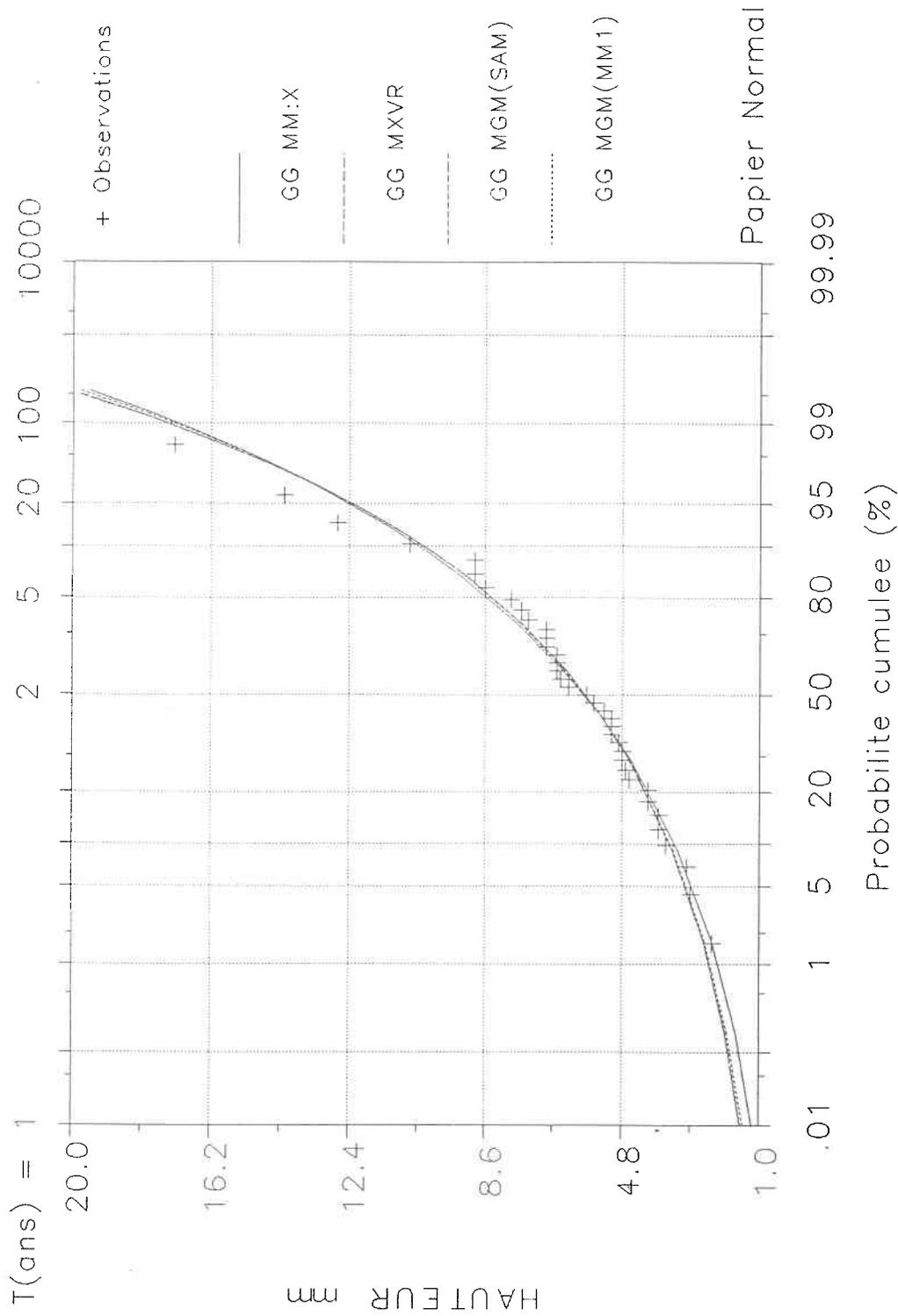


Figure D.9 Ajustement sur papier de probabilité des séries annuelles de maxima - station NB3-5M (GG)

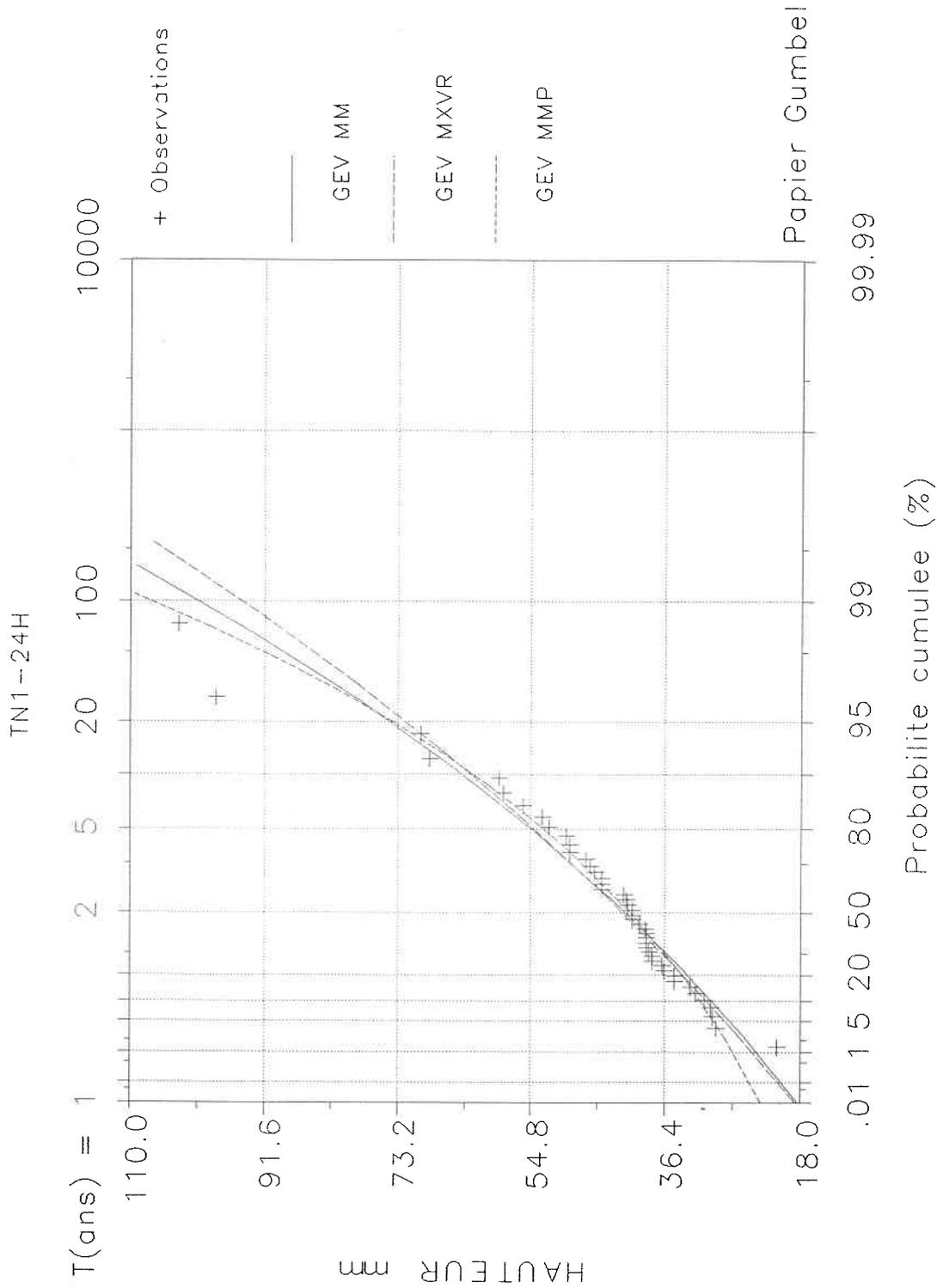


Figure D.10 Ajustement sur papier de probabilité des séries annuelles de maxima - station TN1-24H (GEV)

QU3-24H

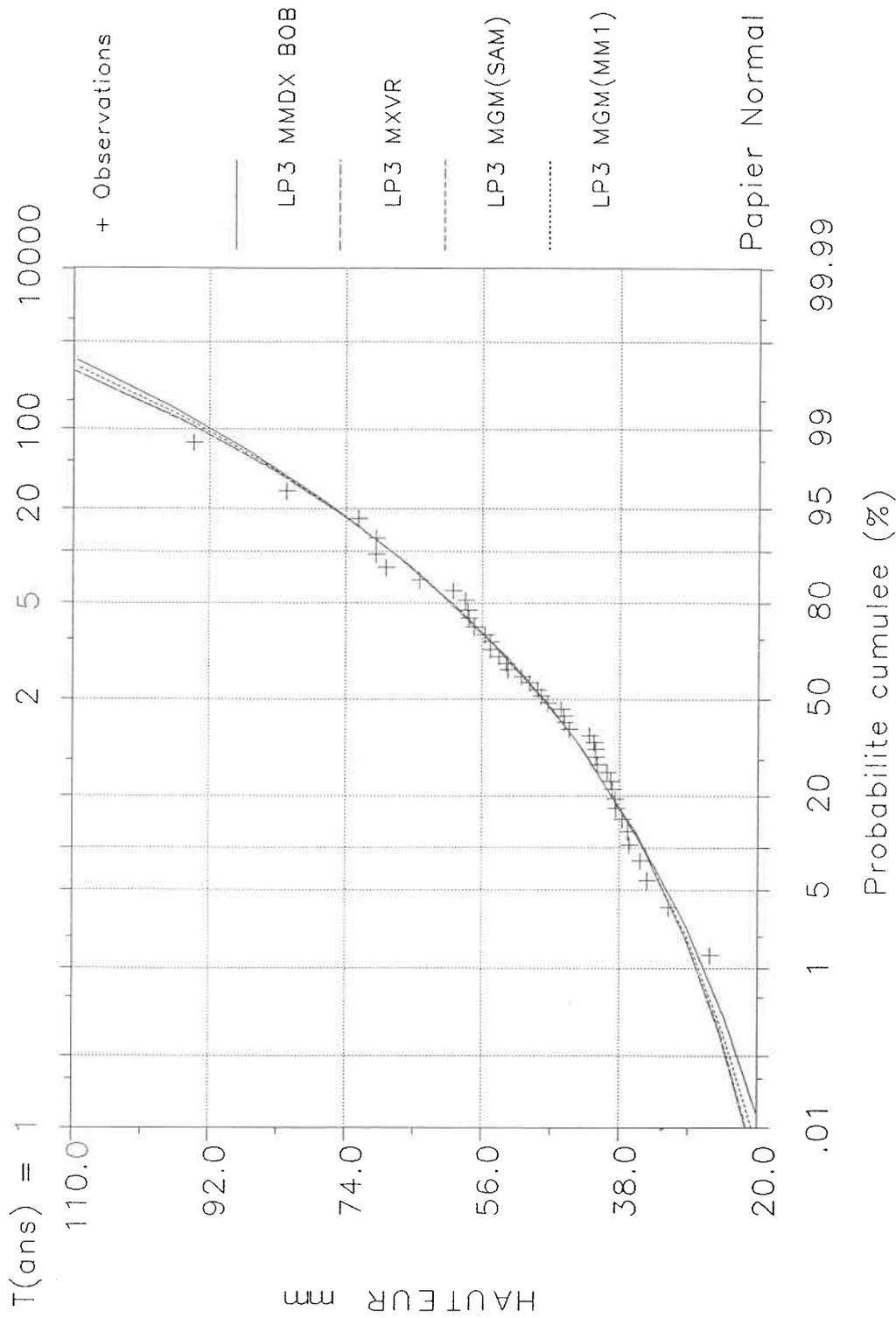


Figure D.11 Ajustement sur papier de probabilité des séries annuelles de maxima - station QU3-24 (LP3)

QU3-24H

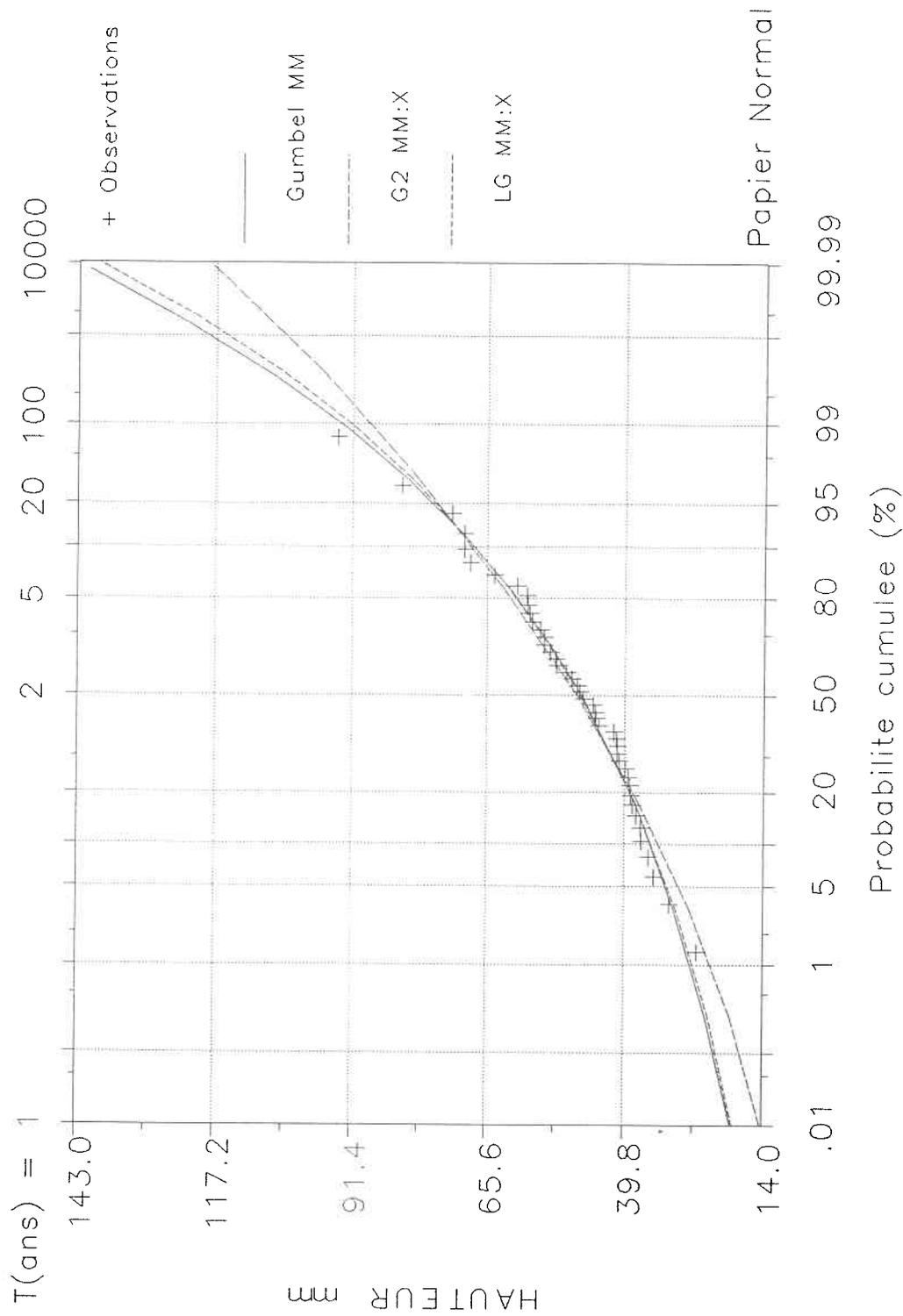


Figure D.12 Ajustement sur papier de probabilité des séries annuelles de maxima - station QU3-24 (GU, GA, LG)

QU3-24H

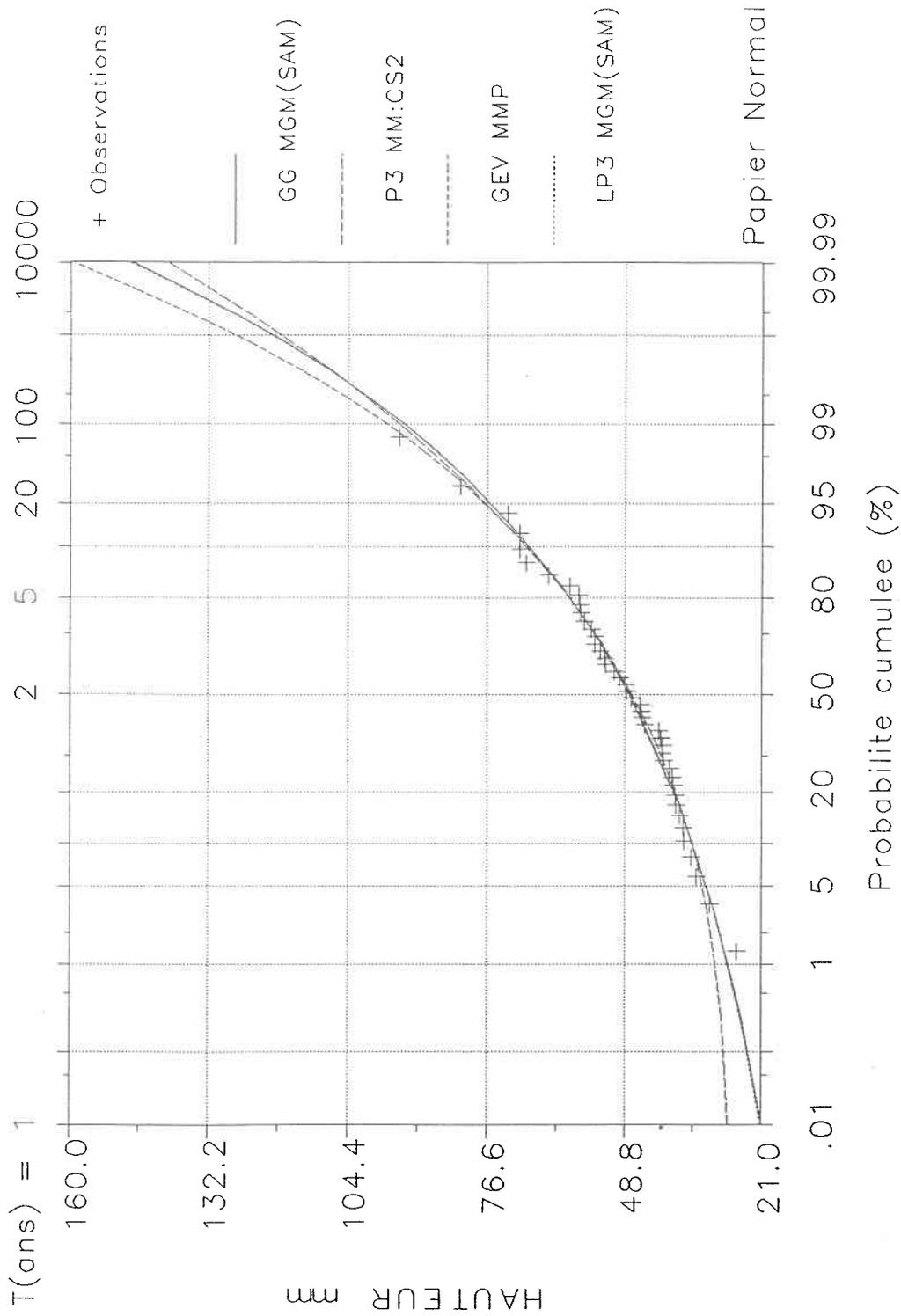


Figure D.13 Ajustement sur papier de probabilité des séries annuelles de maxima - station QU3-24H (GG,P3,GEV,LP3)