

JULIE BÉLIVEAU

**ANALYSE FRÉQUENTIELLE MULTIVARIÉE DE  
LA POINTE, DU VOLUME ET DE LA DURÉE DE  
LA CRUE**

Mémoire présenté  
à la Faculté des études supérieures de l'Université Laval  
dans le cadre du programme de maîtrise en statistique  
pour l'obtention du grade de Maître ès sciences (M.Sc.)

FACULTÉ DES SCIENCES ET DE GÉNIE  
UNIVERSITÉ LAVAL  
QUÉBEC

2006

# Résumé

Les crues sont définies par trois caractéristiques principales, soit la pointe, le volume et la durée. Puisque ces trois variables sont corrélées, trois analyses fréquentielles univariées ne permettent pas d'évaluer complètement la probabilité d'occurrence d'un événement hydrologique. Une classe de copules dites méta-elliptiques, encore peu utilisée en pratique, a donc été employée pour effectuer une modélisation trivariée de la pointe, du volume et de la durée des crues. Un test d'adéquation de type Cramér–von Mises récemment développé par [Genest \*et al.\* \(2006\)](#) a aussi été mis en œuvre afin de trouver le meilleur ajustement. Cette méthodologie a été appliquée sur quatre bassins versants québécois.

# Avant-propos

Je tiens ici à remercier toutes les personnes qui ont contribué, de près ou de loin, à la réalisation de ce mémoire. Merci tout d'abord à ma co-directrice de recherche, Mme Anne-Catherine Favre, professeure au Centre Eau, Terre et Environnement de l'Institut national de la recherche scientifique. Elle a toujours su se montrer disponible et m'a encouragée tout au long de ce projet. J'ai beaucoup apprécié travailler sous sa direction, d'autant plus que ceci m'a permis d'approfondir mes connaissances grâce à ses judicieux conseils.

Je suis également reconnaissante envers mon directeur de recherche, M. Christian Genest, professeur au Département de mathématiques et de statistique de l'Université Laval. En plus de m'avoir proposé ce projet passionnant, il a émis des commentaires constructifs tout au long de sa réalisation.

J'aimerais en outre exprimer toute ma gratitude aux membres de la Chaire de recherche en hydrologie statistique de l'INRS-ETE, avec qui j'ai passé le plus clair de mon temps pendant ma maîtrise. Un merci plus particulier à Mme Christiane Jacques, agente de recherche, qui m'a encouragée, appuyée, conseillée et qui a volontiers répondu à toutes mes questions tout au long de ce travail. Elle m'a été d'un grand soutien, autant du côté moral que technique. Je la remercie également d'avoir accepté de relire mon texte et d'avoir suggéré de nombreuses idées et corrections.

Mon mémoire a aussi bénéficié des commentaires judicieux et approfondis de M. Luc Perreault, chercheur à l'Institut de recherche d'Hydro-Québec à Varennes, qui a aimablement accepté d'agir à titre de rapporteur.

Finalement, je désire bien sûr remercier les membres de ma famille, qui ont toujours encouragé mes études et grâce à qui j'ai grandi dans un milieu épanouissant. Je souhaite également remercier mes amis, qui m'ont fait rire et m'ont permis de me divertir même dans les moments les plus difficiles. Mes derniers remerciements s'adressent à mon amoureux, Éric, qui est toujours un stimulant à aller de l'avant. Il m'a écoutée avec

intérêt parler des hauts et des bas de mes simulations et de ma rédaction, m'a aidée à garder le sourire en temps de stress et a toujours eu confiance en moi. Je le remercie simplement d'avoir été là.

Une partie des travaux décrits dans ce mémoire a été financée par un contrat de recherche entre l'INRS et Hydro-Québec, ainsi que par des subventions de recherche accordées à M. Genest par le Conseil de recherches en sciences naturelles et en génie du Canada et le Fonds québécois de la recherche sur la nature et les technologies. Que ces bailleurs de fonds trouvent ici l'expression de notre reconnaissance.

# Table des matières

Résumé	ii
Avant-Propos	iii
Table des matières	vi
Liste des tableaux	viii
Table des figures	x
<b>1 Introduction</b>	<b>1</b>
<b>2 Analyse univariée</b>	<b>4</b>
2.1 Présentation des données . . . . .	4
2.2 Pointe . . . . .	6
2.2.1 Statistiques descriptives . . . . .	6
2.2.2 Tests d'hypothèse . . . . .	8
2.2.3 Ajustement de lois . . . . .	9
2.3 Volume . . . . .	11
2.3.1 Statistiques descriptives . . . . .	11
2.3.2 Tests d'hypothèse . . . . .	12
2.3.3 Ajustement de lois . . . . .	14
2.4 Durée . . . . .	15
2.4.1 Statistiques descriptives . . . . .	15
2.4.2 Tests d'hypothèse . . . . .	16
2.4.3 Ajustement de lois . . . . .	16
2.5 Synthèse . . . . .	18
<b>3 Analyse de la dépendance</b>	<b>20</b>
3.1 Analyse de la dépendance . . . . .	20
3.2 Synthèse . . . . .	24
<b>4 Aspects théoriques des copules méta-elliptiques</b>	<b>26</b>
4.1 Définition d'une copule . . . . .	26

4.1.1	Copules archimédiennes . . . . .	27
4.1.2	Copules méta-elliptiques . . . . .	28
4.2	Estimation des paramètres d'une copule méta-elliptique . . . . .	29
4.3	Algorithme de simulation d'une copule méta-elliptique . . . . .	32
4.4	Exemples d'applications . . . . .	34
<b>5</b>	<b>Tests d'adéquation</b>	<b>35</b>
5.1	Tests d'adéquation pour les copules . . . . .	35
5.1.1	Algorithme du bootstrap paramétrique à deux niveaux . . . . .	37
5.2	Implantation . . . . .	40
<b>6</b>	<b>Application</b>	<b>41</b>
6.1	Résultats . . . . .	41
<b>7</b>	<b>Conclusion</b>	<b>45</b>
	<b>Bibliographie</b>	<b>47</b>

# Liste des tableaux

2.1	Principales caractéristiques des bassins versants à l'étude . . . . .	5
2.2	Statistiques descriptives de la pointe de l'Ashuapmushuan . . . . .	7
2.3	Tests d'hypothèse pour vérifier le caractère aléatoire de l'échantillon des pointes de l'Ashuapmushuan . . . . .	8
2.4	Paramétrisation de sept lois univariées communément utilisées en hydrologie . . . . .	10
2.5	Critères de comparaison des ajustements des lois de la pointe de l'Ashuapmushuan . . . . .	11
2.6	Statistiques descriptives du volume de l'Ashuapmushuan . . . . .	12
2.7	Tests d'hypothèse pour vérifier le caractère aléatoire de l'échantillon des volumes de l'Ashuapmushuan . . . . .	13
2.8	Critères de comparaison des ajustements des lois du volume de l'Ashuapmushuan . . . . .	14
2.9	Statistiques descriptives de la durée de l'Ashuapmushuan . . . . .	15
2.10	Tests d'hypothèse pour vérifier le caractère aléatoire de l'échantillon des durées de l'Ashuapmushuan . . . . .	16
2.11	Critères de comparaison des ajustements des lois de la durée de l'Ashuapmushuan . . . . .	18
2.12	Tableau synthèse : loi des marginales pour les quatre bassins versants . . . . .	19
3.1	Mesures de la dépendance entre la pointe, le volume et la durée de l'Ashuapmushuan . . . . .	21
3.2	Coefficient de corrélation de Pearson entre la pointe, le volume et la durée pour les quatre bassins versants . . . . .	25
3.3	Rho de Spearman entre la pointe, le volume et la durée pour les quatre bassins versants . . . . .	25
3.4	Tau de Kendall entre la pointe, le volume et la durée pour les quatre bassins versants . . . . .	25
4.1	Fonctions requises pour la simulation de quatre copules méta-elliptiques . . . . .	33

- 6.1 Résultats des ajustements aux copules normale, de Pearson type II, de Student et de Cauchy pour les quatre bassins versants ; les copules choisies sont en gras et les copules rejetées au seuil de 5% sont en italique . 43

# Table des figures

2.1	Localisation des stations à l'étude . . . . .	5
2.2	Représentation graphique des trois variables à l'étude : pointe, volume et durée de crue (hydrogramme de crue de l'Ashuapmushuan pour l'année 1995) . . . . .	6
2.3	Série chronologique de la pointe de l'Ashuapmushuan pour les années 1963 à 1999 . . . . .	7
2.4	Observations sur papier de probabilité normal des quantiles de la pointe en fonction de la loi de probabilité empirique au non-dépassement de l'Ashuapmushuan . . . . .	8
2.5	Graphique des autocorrélations de la variable pointe de l'Ashuapmushuan ; les deux lignes horizontales représentent un intervalle de confiance à 95% . . . . .	9
2.6	Graphique de comparaison des ajustements des lois de la pointe de l'Ashuapmushuan . . . . .	10
2.7	Série chronologique du volume de l'Ashuapmushuan pour les années 1963 à 1999 . . . . .	12
2.8	Observations sur papier de probabilité normal des quantiles du volume en fonction de la loi de probabilité empirique au non-dépassement de l'Ashuapmushuan . . . . .	13
2.9	Graphique des autocorrélations de la variable volume de l'Ashuapmushuan ; les deux lignes horizontales représentent un intervalle de confiance à 95% . . . . .	13
2.10	Graphique de comparaison des ajustements des lois du volume de l'Ashuapmushuan . . . . .	14
2.11	Série chronologique de la durée de l'Ashuapmushuan pour les années 1963 à 1999 . . . . .	15
2.12	Observations sur papier de probabilité normal des quantiles de la durée en fonction de la loi de probabilité empirique au non-dépassement de l'Ashuapmushuan . . . . .	16
2.13	Graphique des autocorrélations de la variable durée de l'Ashuapmushuan ; les deux lignes horizontales représentent un intervalle de confiance à 95% . . . . .	17

2.14	Graphique de comparaison des ajustements des lois de la durée de l'Ashuapmushuan . . . . .	17
2.15	Graphique de comparaison des ajustements des mélanges de lois normale et gamma aux données du volume de la Romaine . . . . .	19
3.1	Graphiques de nuages de points, khi-plot et K-plot illustrant la dépendance entre la pointe, le volume et la durée pour l'Ashuapmushuan . .	22
4.1	Représentation graphique de 10 000 simulations d'une loi de marges $F_1, F_2, F_3$ et de copule méta-elliptique de Student ayant $m$ degrés de liberté ( $m = 2, 3, 4, 5, 6, 7$ ) superposée aux observations de l'Ashuapmushuan . . . . .	31
6.1	Analyse de sensibilité des ajustements des copules de Cauchy et Student, avec paramètre $m = 3$ et 4, selon les données de l'Ashuapmushuan . . .	42
6.2	Représentation graphique de l'ajustement de 10 000 observations de la copule Pearson type II avec paramètre $m = 2$ superposée aux données originales de l'Ashuapmushuan . . . . .	44

# Chapitre 1

## Introduction

Plusieurs applications en hydrologie nécessitent l'utilisation de lois statistiques multivariées (analyse fréquentielle volume-débit de crue, analyse fréquentielle durée-déficit des étiages, modélisation stochastique multi-site des précipitations, ...). Les approches utilisées classiquement sont limitées car elles ne permettent d'utiliser qu'un nombre restreint de distributions (normale multivariée, gamma bivariée, Gumbel bivariée, ...) et dans la majorité des cas en dimension deux. De plus, les marges des lois multivariées classiques doivent être similaires.

Une approche alternative réside dans l'utilisation de copules ([Nelsen, 1999](#)). Les copules sont des fonctions de répartition dont les marges sont uniformes ([Sklar, 1959](#)). Elles permettent de modéliser la structure de dépendance indépendamment des distributions marginales. Grâce à elles, on peut donc former des distributions multidimensionnelles avec différentes marginales et la structure de dépendance sera donnée par la copule.

Dans la littérature, en dehors du domaine de la statistique, c'est surtout en finance et en actuariat que l'on retrouve l'approche par copules, laquelle est encore très peu utilisée dans le domaine de l'hydrologie. On peut néanmoins citer [Favre \*et al.\* \(2004\)](#), [Salvadori et De Michele \(2004\)](#) et [De Michele \*et al.\* \(2005\)](#) qui ont employé les copules pour faire des analyses fréquentielles bivariées.

Les crues sont le plus souvent décrites par trois caractéristiques principales : la pointe, le volume et la durée. L'analyse fréquentielle univariée est l'approche que privilégient généralement les hydrologues pour analyser le risque associé aux événements hydrologiques extrêmes. Cette méthode statistique consiste à étudier des événements historiques afin d'en évaluer les probabilités d'apparition future. Elle permet de four-

nir aux autorités responsables de l'aménagement les éléments de décision nécessaires. De plus, elle livre aux ingénieurs chargés de la réalisation des ouvrages les bases de dimensionnement nécessaires.

Cependant, comme ces trois variables sont corrélées, des analyses fréquentielles univariées ne sont pas en mesure de fournir une évaluation complète de la probabilité d'occurrence de l'événement considéré. Prenons par exemple la crue de conception d'un ouvrage hydraulique en rivière, laquelle est déterminée par la pointe et le volume de la crue. Pour ces deux variables, on doit obtenir des valeurs correspondant à une probabilité conjointe au non-dépassement  $p$ , ou de façon équivalente à une période de retour bivariée  $T = 1/(1 - p)$ . Cette estimation doit être aussi précise que possible. En effet, une surestimation conduit à un sur-dimensionnement des ouvrages, entraînant ainsi des coûts de construction trop élevés, alors qu'une sous-estimation augmente les risques d'inondation ou de bris de l'ouvrage, et éventuellement des pertes matérielles, voire humaines.

Plusieurs études considèrent une modélisation bivariée, soit du débit et du volume, soit du volume et de la durée; voir par exemple [Singh et Singh \(1991\)](#); [Goel \*et al.\* \(1998\)](#); [Yue \(2000\)](#) pour des modélisations débit—volume et [Yue \*et al.\* \(1999\)](#) pour un exemple de modélisation volume—durée. Cependant aucune étude n'a encore été effectuée en dimension trois sans utiliser une hypothèse simplificatrice de normalité qui est souvent erronée.

L'objectif de ce mémoire est de réaliser une analyse fréquentielle trivariée des crues (pointe, volume et durée) en utilisant les copules. Ainsi, les résultats de cette analyse permettront de produire par exemple la distribution conjointe associée aux données. Plus particulièrement, la classe des copules méta-elliptiques sera étudiée, compte tenu de sa flexibilité et du fait qu'elle peut être employée en dimension  $p \geq 2$  quelconque. De plus, un test d'adéquation basé sur la statistique de Cramér—von Mises ([Genest \*et al.\*, 2006](#)) justifiera le choix de la copule la mieux adaptée aux données. Pour ce faire, l'utilisation d'un bootstrap paramétrique à deux niveaux nous permettra de comparer la distribution théorique à la distribution échantillonnale et ainsi d'obtenir un seuil observé.

Le chapitre II introduit d'abord les données utilisées tout au long de ce mémoire. Il expose également une analyse univariée sur chacune des variables. Le chapitre III présente une analyse de dépendance pour chaque paire de variables étudiées. Quant au chapitre IV, il définit plus précisément la notion de copule méta-elliptique ainsi que la façon de la simuler. Le chapitre V présente les tests d'adéquation des copules, l'algorithme du bootstrap paramétrique à deux niveaux ainsi que quelques réflexions sur

l'implantation informatique de ces algorithmes. Pour finir, le chapitre VI dévoile les résultats des applications effectuées selon les méthodes proposées aux chapitres précédents. Une courte conclusion complète l'ouvrage.

# Chapitre 2

## Analyse univariée

### 2.1 Présentation des données

L'étude porte sur quatre bassins versants, à savoir ceux des rivières Ashuapmushuan, Batiscan, Harricana et Romaine. La figure 2.1 montre la localisation exacte des cours d'eau correspondants sur le territoire québécois. À première vue, on remarque qu'elles sont situées dans quatre régions administratives différentes. D'ailleurs, ces rivières ont été choisies pour leurs caractéristiques distinctes, telles que présentées au tableau 2.1, dont la taille du bassin versant et le nombre d'années d'observations disponibles. Dans deux cas, les données ne représentent pas des séries continues, du fait que la collecte a été interrompue pendant un certain laps de temps.

Pour chacune des années mentionnées dans le tableau 2.1, on dispose de séries d'observations journalières de débit (en  $\text{m}^3/\text{s}$ ) de chacune des rivières. Le débit est le volume d'eau passant en un point d'un cours d'eau par unité de temps, ce qui est généralement mesuré en mètres cubes par seconde. Pour déterminer le débit moyen d'une rivière, on prend à plusieurs endroits des mesures de profondeur du cours d'eau, ainsi que de la vitesse du courant, puis on intègre le tout sur la section de la rivière. À partir de cela, on peut donc construire un graphique des débits journaliers de la rivière, qu'on nomme un hydrogramme.

La crue est la hausse du niveau de l'eau, et donc l'augmentation du débit d'un cours d'eau. Au Québec, on distingue généralement deux événements de crue par année : la crue de printemps et la crue d'été/automne. La première est conditionnée par la fonte de la neige et les précipitations liquides du printemps. La seconde est induite unique-

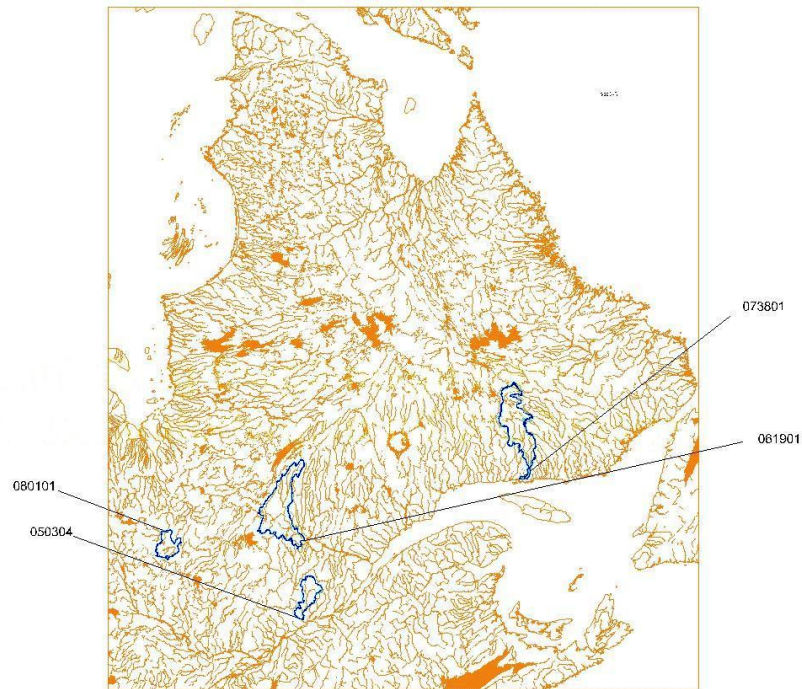


FIG. 2.1 – Localisation des stations à l'étude

Rivière	Région administrative	Taille du bassin (km <sup>2</sup> )	Période d'échantillonnage (années)	Étendue finale
Ashuapmushuan 061901	Saguenay-Lac-Saint-Jean	15 300	37	1963 à 1999
Batiscan 050304	Mauricie	4 480	52	1924 à 1942 1968 à 2000
Harricana 080101	Abitibi-Témiscamingue	3 680	85	1915 à 1999
Romaine 073801	Côte-Nord	13 000	47	1957 à 1959 1961 à 2004

TAB. 2.1 – Principales caractéristiques des bassins versants à l'étude

ment par de fortes précipitations liquides. Dans le présent mémoire, on a considéré uniquement les crues printanières. Il est à noter que celles-ci correspondent aux crues maximales annuelles sur les grands bassins versants québécois.

La durée de la crue est égale au nombre de jours entre le début et la fin de la crue, tel qu'évalué par des experts hydrologues, à partir d'un hydrogramme. Le volume est

l'aire sous la courbe située entre le début et la fin de la crue, telle qu'on peut l'observer dans un hydrogramme. La pointe est définie comme étant le débit maximal survenant entre les début et fin de crue identifiés par les experts. Les définitions de la crue, du volume et de la durée sont illustrées à la figure 2.2.

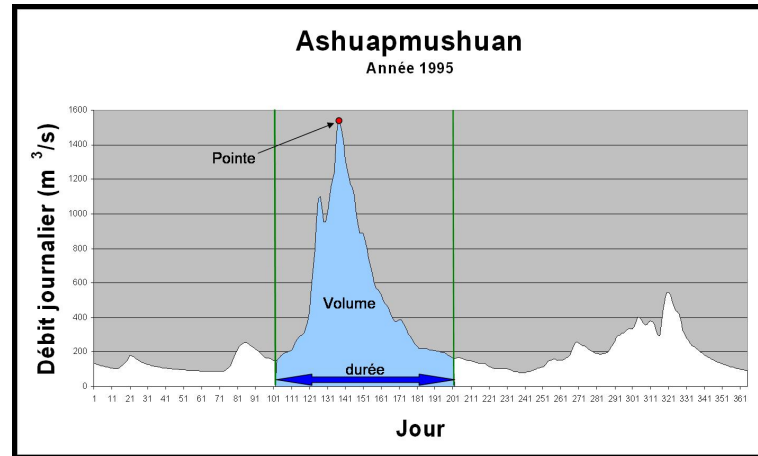


FIG. 2.2 – Représentation graphique des trois variables à l'étude : pointe, volume et durée de crue (hydrogramme de crue de l'Ashuapmushuan pour l'année 1995)

Le présent chapitre propose un exemple d'analyse fréquentielle univariée de chacune des trois variables, soit la pointe, le volume et la durée, pour l'Ashuapmushuan. On cherche à étudier les événements passés sur ce bassin versant dans le but de prédire les probabilités d'apparition future. Pour ce qui est des résultats concernant les autres rivières, il est possible de se référer à Jacques (2004).

## 2.2 Pointe

### 2.2.1 Statistiques descriptives

La figure 2.3 illustre le graphique d'une série temporelle de la pointe en fonction de l'année. Elle ne suggère la présence d'aucune donnée aberrante ou de tendance dans les observations.

Une analyse exploratoire a ensuite été effectuée pour avoir une idée de la moyenne, de l'écart-type, du minimum, du maximum, de la médiane et de divers coefficients concernant les pointes sur le bassin versant de l'Ashuapmushuan. Le tableau 2.2 présente ces

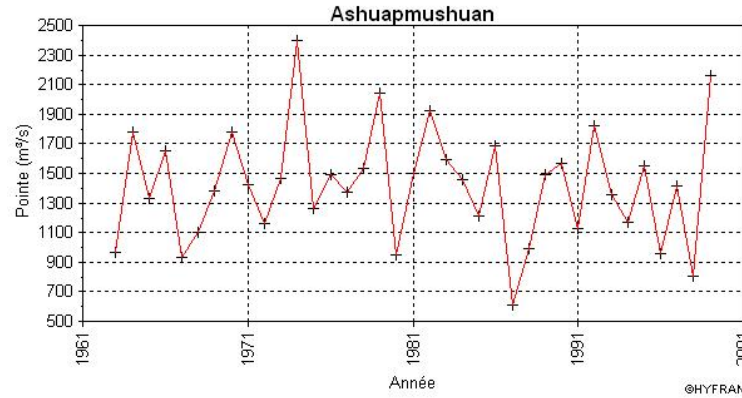


FIG. 2.3 – Série chronologique de la pointe de l’Ashuapmushuan pour les années 1963 à 1999

résultats. Les coefficients d’asymétrie et d’aplatissement indiquent que la distribution est légèrement asymétrique et plus aplatie qu’une distribution normale.

Statistique	
Taille de l’échantillon	37
Débit minimal	610
Débit maximal	2400
Débit moyen	1420
Écart-type	383
Débit médian	1420
Coefficient de variation	0,27
Coefficient d’asymétrie	0,32
Coefficient d’aplatissement	2,93

TAB. 2.2 – Statistiques descriptives de la pointe de l’Ashuapmushuan

Par la suite, une vérification de la normalité a été réalisée à l’aide d’un graphique et des tests classiques de Shapiro–Wilk, Kolmogorov–Smirnov, Cramér–von Mises et Anderson–Darling. Aucun des tests ne permet de réfuter l’hypothèse selon laquelle l’échantillon de la pointe est normalement distribué. Ceci est confirmé visuellement par la droite de Henry présentée à la figure 2.4. Sur ce graphique, communément appelé papier de probabilité normal par les hydrologues (voir Haché *et al.* (1999)), les observations sont tracées en fonction des probabilités empiriques de l’échantillon. Pour un échantillon normal, les observations devraient être alignées, ce qui est bien le cas ici.

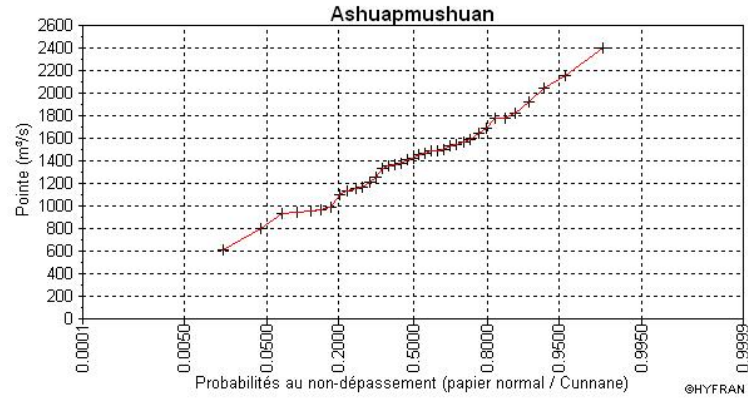


FIG. 2.4 – Observations sur papier de probabilité normale des quantiles de la pointe en fonction de la loi de probabilité empirique au non-dépassement de l’Ashuapmushuan

## 2.2.2 Tests d’hypothèse

Le tableau 2.3 présente les résultats de trois tests utilisés pour vérifier si les données de l’échantillon sont indépendantes et identiquement distribuées. On doit s’assurer de cette condition afin de pouvoir ajuster une distribution statistique par la suite.

Tests d’hypothèse	Statistique	Seuil observé
Test d’indépendance de Wald–Wolfowitz	$ U  = 1,55$	0,1221
Test de stationnarité de Kendall	$ K  = 0,04$	0,9687
Test d’homogénéité de Wilcoxon	$ W  = 0,15$	0,8792

TAB. 2.3 – Tests d’hypothèse pour vérifier le caractère aléatoire de l’échantillon des pointes de l’Ashuapmushuan

Le test de Wald–Wolfowitz permet de vérifier l’indépendance mutuelle des données. Son seuil étant supérieur à 5%, on peut conclure qu’il n’y a pas d’autocorrélation entre les observations. De plus, le graphique 2.5 des autocorrélations de la variable pointe appuie cette affirmation.

Le test de Kendall permet d’éprouver l’hypothèse selon laquelle les données sont stationnaires, c’est-à-dire que la série n’a pas tendance à augmenter (ou diminuer) avec le temps. Cette hypothèse nulle étant ici acceptée au seuil de 5%, on conclut qu’il n’y a aucune tendance dans les données.

Le test d’homogénéité de Wilcoxon permet de vérifier si la moyenne de la série de

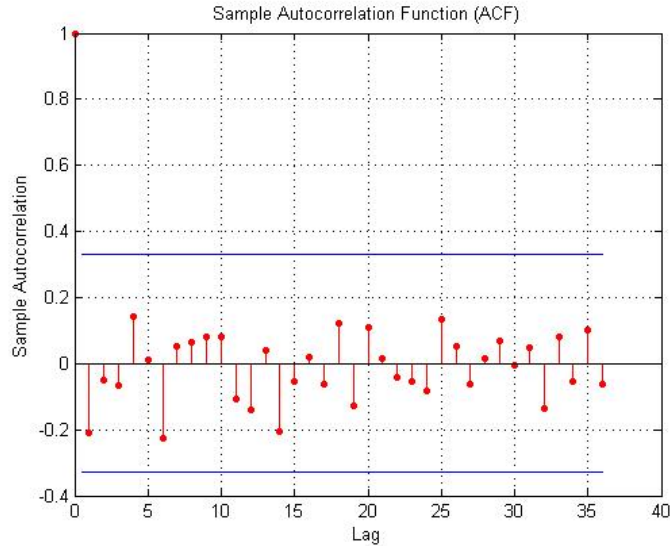


FIG. 2.5 – Graphique des autocorrélations de la variable pointe de l’Ashuapmushuan ; les deux lignes horizontales représentent un intervalle de confiance à 95%

1963 à 1981 est significativement différente de celle de la série de 1982 à 1999. Le seuil observé très fort nous conduit à accepter l’hypothèse nulle à l’effet que les moyennes des deux sous-échantillons sont égales.

### 2.2.3 Ajustement de lois

Bien que la pointe de l’Ashuapmushuan puisse être modélisée à l’aide de la loi normale, plusieurs autres distributions statistiques ont été envisagées. Pour ce faire, nous avons utilisé le logiciel Hyfran ([Chaire en hydrologie statistique, 2002](#)), qui permet d’ajuster les 15 lois de probabilité les plus utilisées en hydrologie. D’un point de vue théorique, la loi GEV devrait être utilisée, mais comme la taille de l’échantillon est faible, il est possible que d’autres lois conviennent mieux aux données. Le tableau 2.4 précise la paramétrisation des lois utilisées dans la suite.

Un graphique superposant les probabilités empiriques et théoriques des distributions gamma, normale, GEV, Pearson type III et log-Pearson type III a été réalisé. Ce graphique est présenté à la figure 2.6.

Le tableau 2.5 présente en outre la valeur observée des critères d’information bayésien (BIC) et de Akaike (AIC). Ces critères permettent de comparer le degré d’ajuste-

$f(x)$	loi
$f(x) = \frac{1}{\alpha} \left\{ 1 - \frac{k}{\alpha} (x - u) \right\}^{\frac{1}{k}-1} \exp \left[ - \left\{ 1 - \frac{k}{\alpha} (x - u) \right\}^{\frac{1}{k}} \right]$	GEV
$f(x) = \frac{1}{\alpha} \exp \left\{ - \frac{x-u}{\alpha} - \exp \left( - \frac{x-u}{\alpha} \right) \right\}$	Gumbel
$f(x) = \frac{\alpha^\lambda}{\Gamma(\lambda)} (x - m)^{\lambda-1} \exp \{ -\alpha(x - m) \}$	Pearson type III
$f(x) = \frac{\alpha^\lambda}{\Gamma(\lambda)x} (\ln x - m)^{\lambda-1} \exp \{ -\alpha(\ln x - m) \}$	log-Pearson type III
$f(x) = \frac{\alpha^\lambda}{\Gamma(\lambda)} x^{\lambda-1} \exp(-\alpha x)$	gamma
$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp \left\{ - \frac{(x-\mu)^2}{2\sigma^2} \right\}$	normale
$f(x) = \frac{1}{\sqrt{2\pi}\sigma x} \exp \left\{ - \frac{(\ln x - \mu)^2}{2\sigma^2} \right\}$	log-normale
$f(x) = \Pi \frac{\alpha_1^{\lambda_1}}{\Gamma(\lambda_1)} x^{\lambda_1-1} \exp(-\alpha_1 x) + (1 - \Pi) \frac{\alpha_2^{\lambda_2}}{\Gamma(\lambda_2)} x^{\lambda_2-1} \exp(-\alpha_2 x)$	mélange de gamma
$f(x) = \Pi \frac{1}{\sqrt{2\pi}\sigma_1} \exp \left\{ - \frac{(x-\mu_1)^2}{2\sigma_1^2} \right\} + (1 - \Pi) \frac{1}{\sqrt{2\pi}\sigma_2} \exp \left\{ - \frac{(x-\mu_2)^2}{2\sigma_2^2} \right\}$	mélange de normales

TAB. 2.4 – Paramétrisation de sept lois univariées communément utilisées en hydrologie

ment des différentes lois, en tenant compte du nombre de leurs paramètres. Les meilleurs ajustements correspondent aux plus faibles valeurs de ces statistiques. On peut se référer à [Press \(2003\)](#) pour avoir une définition plus détaillée de ces deux critères.

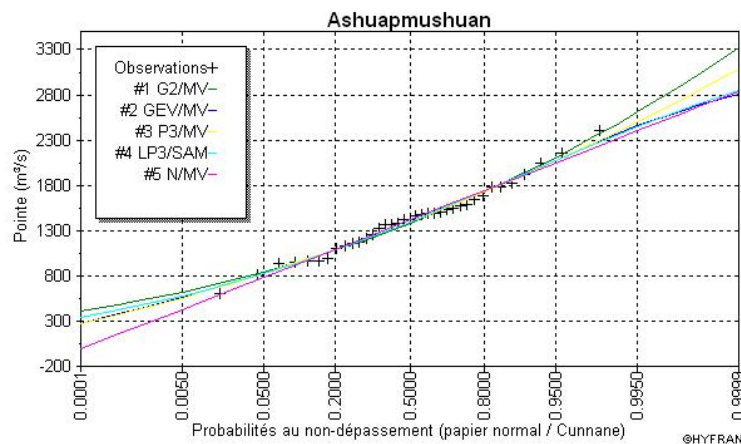


FIG. 2.6 – Graphique de comparaison des ajustements des lois de la pointe de l'Ashuapmushuan

Comme le but de l'analyse fréquentielle est d'estimer le risque associé aux événements hydrologiques extrêmes et donc d'extrapoler les quantiles, nous avons privilégié les lois qui ajustaient le mieux les observations extrêmes (queue supérieure de la distribution) visuellement. Nous avons donc conclu que la loi gamma de paramètres  $\hat{\alpha} = 0,01$

Modèle	Nb	BIC	AIC
Loi gamma (Maximum de vraisemblance)	2	551,12	547,90
Loi normale (Maximum de vraisemblance)	2	551,37	548,15
Loi GEV (Maximum de vraisemblance)	3	554,37	549,54
Loi Pearson type III (Maximum de vraisemblance)	3	554,37	549,54
Loi log-Pearson type III (Méthode SAM)	3	554,40	549,57

TAB. 2.5 – Critères de comparaison des ajustements des lois de la pointe de l’Ashuap-mushuan

et  $\hat{\lambda} = 13,45$  (courbe en vert foncé dans la figure 2.6) représente le meilleur compromis pour l’ajustement de la variable pointe, particulièrement à des fins d’extrapolation. De plus, dans ce cas les critères BIC et AIC sont minimums. Les estimations des paramètres sont celles produites par la méthode du maximum de vraisemblance.

## 2.3 Volume

### 2.3.1 Statistiques descriptives

La figure 2.7 présente le graphique de la série temporelle du volume (en hm<sup>3</sup>) en fonction des années. À première vue, rien d’inhabituel ne ressort de ce graphique.

Le tableau 2.6 rapporte les statistiques descriptives calculées sur les données dans le but de quantifier leur variation. On remarque à nouveau une très légère asymétrie de la distribution, comme celle qui a déjà été observée pour la pointe. De plus, le coefficient d’aplatissement est un peu moins élevé que celui d’une loi normale.

L’hypothèse de normalité a été vérifiée par les mêmes méthodes que pour la pointe. D’après la figure 2.8, les observations en fonction des probabilités empiriques sont relativement bien alignées. L’hypothèse a donc été retenue, d’autant plus qu’aucun des quatre tests ne permettait de la refuser au seuil de 5%.

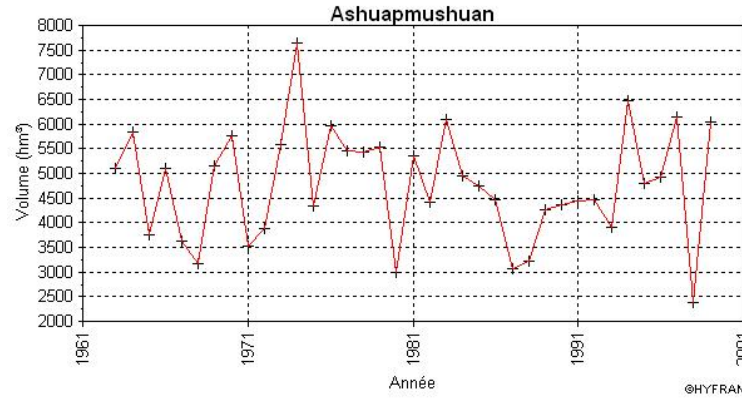


FIG. 2.7 – Série chronologique du volume de l’Ashuapmushuan pour les années 1963 à 1999

Statistique	
Taille de l’échantillon	37
Volume minimal	2369
Volume maximal	7637
Volume moyen	4767
Écart-type	1133
Volume médian	4801
Coefficient de variation	0,24
Coefficient d’asymétrie	0,07
Coefficient d’aplatissement	2,65

TAB. 2.6 – Statistiques descriptives du volume de l’Ashuapmushuan

### 2.3.2 Tests d’hypothèse

Le tableau 2.7 donne les résultats des trois tests utilisés en vue de vérifier si les données de l’échantillon sont indépendantes et identiquement distribuées. Comme on peut le constater, rien ne permet de rejeter cette hypothèse, au seuil de 5%. De plus, le graphique des autocorrélations, présenté à la figure 2.9, nous indique aucune autocorrélation significative entre les données.

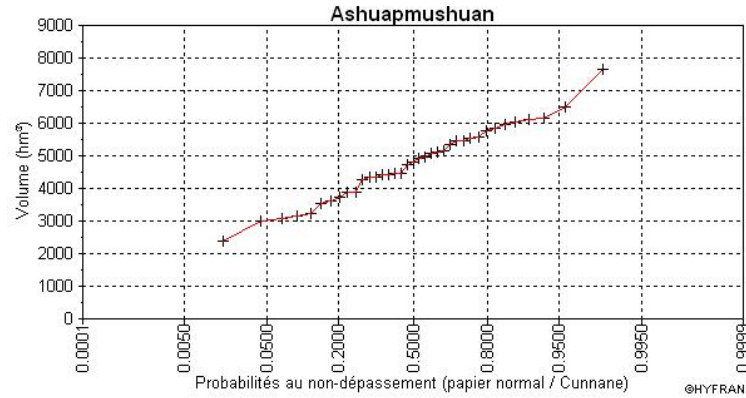


FIG. 2.8 – Observations sur papier de probabilité normal des quantiles du volume en fonction de la loi de probabilité empirique au non-dépassement de l’Ashuapmushuan

Tests d’hypothèse	Statistique	Seuil observé
Test d’indépendance de Wald–Wolfowitz	$ U  = 0,33$	0,7407
Test de stationnarité de Kendall	$ K  = 0,09$	0,9270
Test d’homogénéité de Wilcoxon	$ W  = 0,62$	0,5334

TAB. 2.7 – Tests d’hypothèse pour vérifier le caractère aléatoire de l’échantillon des volumes de l’Ashuapmushuan

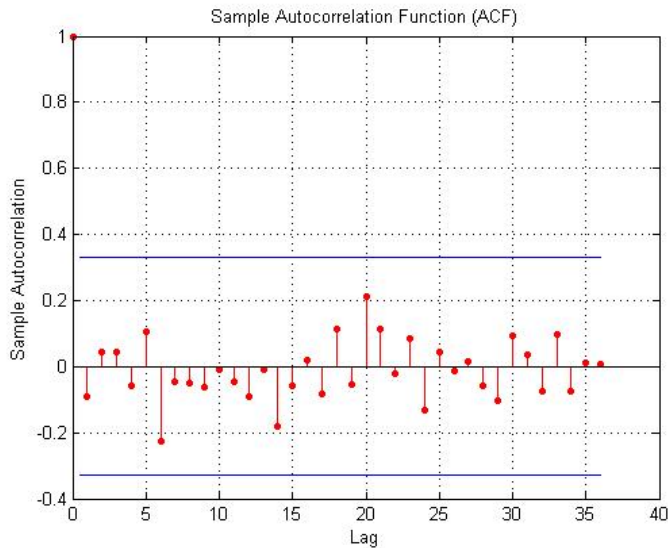


FIG. 2.9 – Graphique des autocorrélations de la variable volume de l’Ashuapmushuan ; les deux lignes horizontales représentent un intervalle de confiance à 95%

### 2.3.3 Ajustement de lois

La figure 2.10 présente un graphique sur papier de probabilité normal, superposant les observations aux lois normale, log-Pearson type III, Pearson type III et GEV. De plus, le tableau 2.8 sert à comparer numériquement les ajustements des différentes lois. Comme l'ajustement visuel est très semblable pour les différentes lois, nous avons choisi la loi présentant le meilleur critère BIC et AIC. La loi normale de paramètres  $\hat{\mu} = 4766,95$  et  $\hat{\sigma} = 1133,06$  était donc la plus appropriée pour les données. Les estimations de ses paramètres ont été obtenues par la méthode du maximum de vraisemblance.

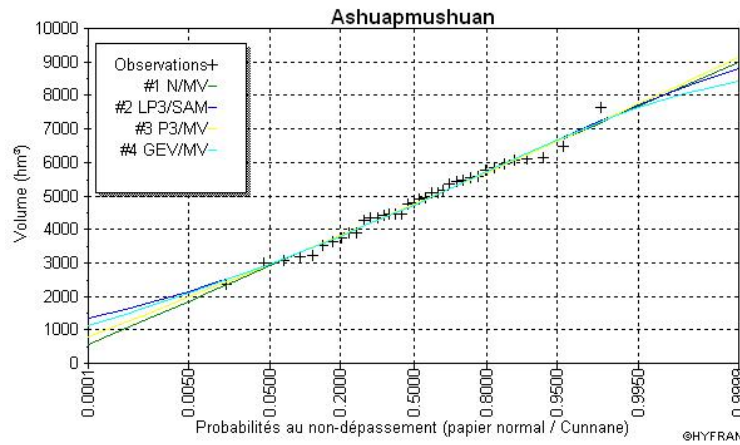


FIG. 2.10 – Graphique de comparaison des ajustements des lois du volume de l'Ashuapmushuan

Modèle	Nb	BIC	AIC
Loi normale (Maximum de vraisemblance)	2	631,64	628,42
Loi log-Pearson type III (Maximum SAM)	3	635,07	630,24
Loi Pearson type III (Maximum de vraisemblance)	3	635,20	630,37
Loi GEV (Maximum de vraisemblance)	3	635,25	630,41

TAB. 2.8 – Critères de comparaison des ajustements des lois du volume de l'Ashuapmushuan

## 2.4 Durée

### 2.4.1 Statistiques descriptives

La figure 2.11 illustre la série chronologique représentant la durée (en jours) de la crue sur l'Ashuapmushuan. Les statistiques descriptives de l'échantillon sont également présentées au tableau 2.9.

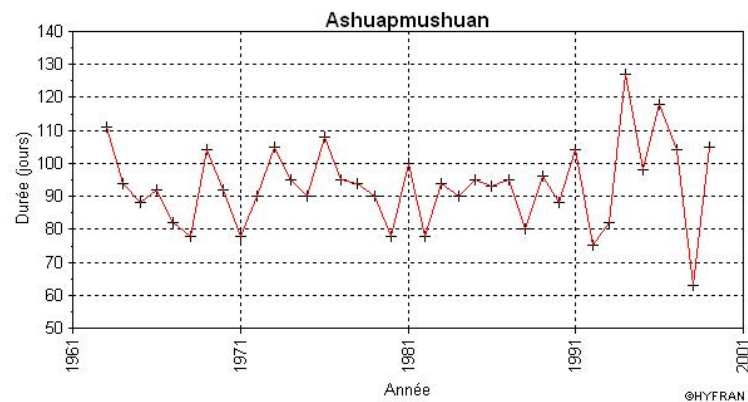


FIG. 2.11 – Série chronologique de la durée de l'Ashuapmushuan pour les années 1963 à 1999

Statistique	
Taille de l'échantillon	37
Durée minimale	63
Durée maximale	127
Durée moyenne	93,22
Écart-type	12,59
Durée médiane	94
Coefficient de variation	0,14
Coefficient d'asymétrie	0,25
Coefficient d'aplatissement	3,37

TAB. 2.9 – Statistiques descriptives de la durée de l'Ashuapmushuan

D'après la droite de Henry présentée à la figure 2.12, les observations sont bien alignées, à l'exception des valeurs extrêmes. On peut donc supposer ici encore que les données sont normales. D'ailleurs, ce choix est cohérent avec les seuils observés des quatre tests de normalité, qui sont tous supérieurs à 15%.

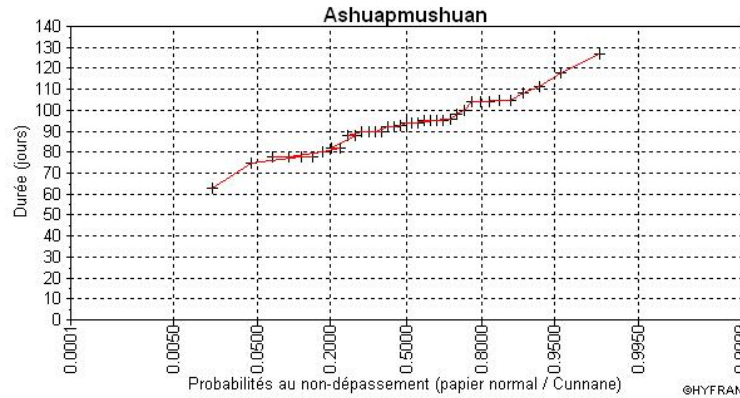


FIG. 2.12 – Observations sur papier de probabilité normal des quantiles de la durée en fonction de la loi de probabilité empirique au non-dépassement de l'Ashuapmushuan

## 2.4.2 Tests d'hypothèse

Les tests d'indépendance, de stationnarité et d'homogénéité nous indiquent que les données sont indépendantes et identiquement distribuées. En effet, l'hypothèse nulle est acceptée au seuil de 5% dans les trois cas ; voir le tableau 2.10. Le graphique 2.13 confirme également l'hypothèse d'indépendance.

Tests d'hypothèse	Statistique	Seuil observé
Test d'indépendance de Wald–Wolfowitz	$ U  = 0,44$	0,6562
Test de stationnarité de Kendall	$ K  = 0,75$	0,4560
Test d'homogénéité de Wilcoxon	$ W  = 0,27$	0,7840

TAB. 2.10 – Tests d'hypothèse pour vérifier le caractère aléatoire de l'échantillon des durées de l'Ashuapmushuan

## 2.4.3 Ajustement de lois

Selon la figure 2.14, les lois normale, log-normale, Pearson type III et GEV s'ajustent assez similairement aux observations. Cependant, la loi log-normale semble mieux s'ajus-

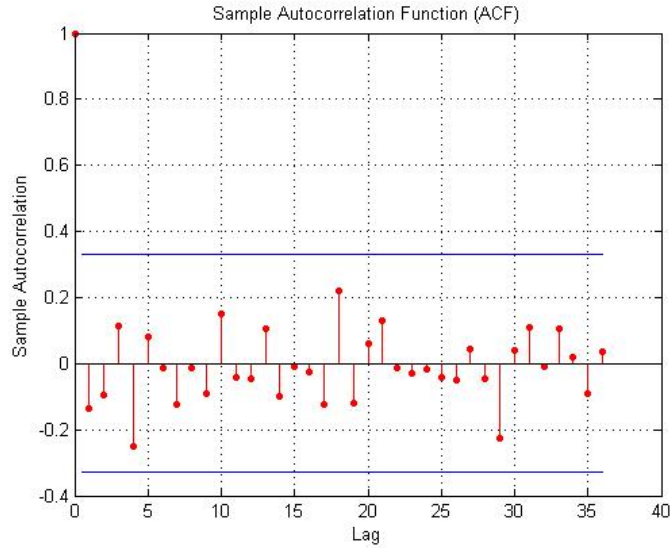


FIG. 2.13 – Graphique des autocorrélations de la variable durée de l’Ashuapmushuan ; les deux lignes horizontales représentent un intervalle de confiance à 95%

ter à la queue supérieure de la distribution. Le tableau 2.11 donne aussi d’autres caractéristiques sur l’ajustement des lois. Il est à souligner que les critères BIC et AIC sont quasiment équivalents pour les lois normale et log-normale. Nous avons par conséquent choisi de privilégier la loi log-normale de paramètres  $\hat{\mu} = 4,53$  et  $\hat{\sigma} = 0,14$ . Il s’agit, comme précédemment, d’estimations à vraisemblance maximale.

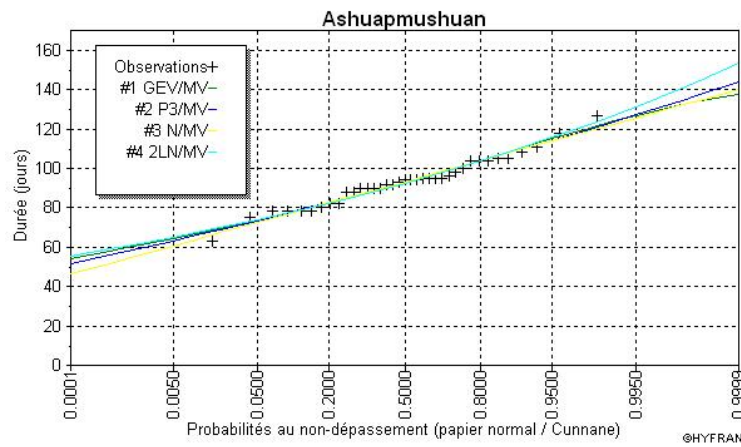


FIG. 2.14 – Graphique de comparaison des ajustements des lois de la durée de l’Ashuapmushuan

Modèle	$Nb$	BIC	AIC
Loi normale (Maximum de vraisemblance)	2	298,65	295,42
Loi log-normale (Maximum de vraisemblance)	2	298,77	295,55
Loi Pearson type III (Maximum de vraisemblance)	3	301,99	297,15
Loi GEV (Maximum de vraisemblance)	3	302,37	297,54

TAB. 2.11 – Critères de comparaison des ajustements des lois de la durée de l’Ashuap-mushuan

## 2.5 Synthèse

Le tableau 2.12 présente un résumé des résultats pour les quatre bassins versants à l’étude. En effet, une analyse fréquentielle similaire à celle présentée dans ces dernières pages a également été effectuée pour les trois autres rivières.

Il est à souligner que des hypothèses d’homogénéité et de stationnarité ont dû être supposées pour le cas de la rivière Batiscan, étant donné qu’aucune mesure de débit ne fut prise entre les années 1943 et 1967.

Une difficulté particulière a été rencontrée dans le cas de la Romaine, puisque le test de Wilcoxon a alors conduit au rejet de l’hypothèse nulle d’homogénéité des séries dans les cas des variables pointe et volume. On a alors dû trouver un mélange de deux lois s’adaptant bien à la situation. Un mélange de lois comprend les paramètres associés à chacune des deux lois utilisées, en plus d’un paramètre représentant la probabilité de choisir la première des deux lois.

Dans le cas du volume de la Romaine, par exemple, on a essayé deux mélanges de lois, soit en utilisant deux lois normales ou encore deux lois gammas. La figure 2.15 présente les résultats de ces deux essais. Comme on peut le voir, le mélange de lois normales s’ajuste mieux aux données du volume de la Romaine.

Rivière	Pointe	Volume	Durée
Ashuapmushuan	gamma (0,01 ; 13,45)	normale (4766,95 ; 1133,06)	log-normale (4,53 ; 0,14)
Batiscan	GEV (142,93 ; 0,23 ; 534,88)	Pearson type III (0,03 ; 79,33 ; -1471,06)	Pearson type III (0,32 ; 20,72 ; 26,22)
Harricana	log-normale (5,20 ; 0,25)	Pearson type III (0,01 ; 6,76 ; 408,52)	Gumbel (9,95 ; 107,69)
Romaine	Mélange de gamma (0,23 ; 15,86 ; 0,01 ; 3,85 ; 0,01)	Mélange de normale (0,60 ; 4478,51 ; 823,16 ; 5719,20 ; 489,32)	gamma (0,45 ; 41,63)

TAB. 2.12 – Tableau synthèse : loi des marginales pour les quatre bassins versants

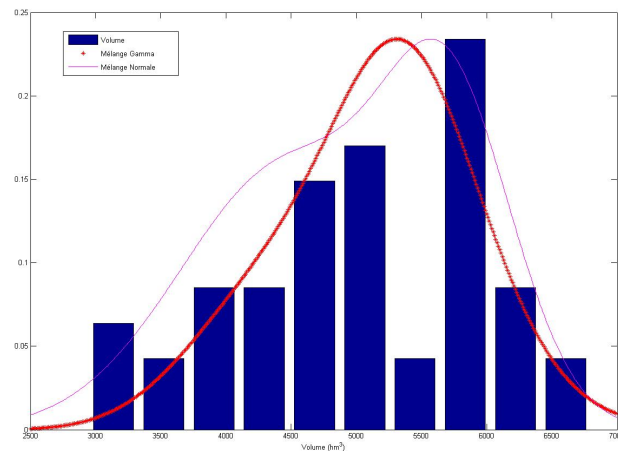


FIG. 2.15 – Graphique de comparaison des ajustements des mélanges de lois normale et gamma aux données du volume de la Romaine

# Chapitre 3

## Analyse de la dépendance

### 3.1 Analyse de la dépendance

L'analyse de la dépendance peut démarrer de deux manières différentes, soit par l'examen de caractéristiques statistiques ou encore par des graphiques. Au plan statistique, trois mesures classiques ont été utilisées, à savoir le coefficient de corrélation de Pearson, la corrélation des rangs de Spearman et le tau de Kendall. Leurs définitions, rappelées ci-dessous, sont tirées de [Capéraà et Van Cutsem \(1988\)](#).

Le coefficient de corrélation de Pearson, défini par

$$r = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2 \sum_{i=1}^n (Y_i - \bar{Y})^2}}, \quad (3.1)$$

est une mesure d'association paramétrique entre deux variables aléatoires continues. Il mesure l'existence d'une relation linéaire entre les données.

Le rho de Spearman et le tau de Kendall sont pour leur part des mesures d'association non paramétriques basées sur les rangs des observations. Le rho de Spearman est défini par

$$\rho_n = 1 - \frac{6}{n(n^2 - 1)} \sum_{i=1}^n (S_i - R_i)^2, \quad (3.2)$$

où  $S_i$  et  $R_i$  sont respectivement les rangs des variables  $X_i$  et  $Y_i$ .

Quant au tau de Kendall, il est donné par

$$\tau_n = \frac{2(C - D)}{n(n - 1)}, \quad (3.3)$$

où  $C$  et  $D$  représentent respectivement le nombre de paires concordantes et discordantes. On dit que deux paires  $(X_i, Y_i)$  et  $(X_j, Y_j)$  sont concordantes lorsque

$$(X_i - X_j)(Y_i - Y_j) > 0$$

et qu'elles sont discordantes quand

$$(X_i - X_j)(Y_i - Y_j) < 0.$$

Noter que la probabilité d'avoir  $(X_i - X_j)(Y_i - Y_j) = 0$  est nulle lorsque les lois des variables concernées sont continues.

Dans les trois cas, les équations présentées sont les versions échantillonnales des coefficients. Il existe aussi une valeur théorique pour chacun de ceux-ci, obtenue à partir de la population. On remarque que si les variables sont indépendantes, les valeurs théoriques de chacun de ces coefficients seront égales à 0 et leurs valeurs expérimentales seront généralement proches de 0. Par ailleurs, une corrélation théorique ou échantillonnale près de 1 ou de  $-1$  indique une dépendance très forte, qui est soit positive ou négative selon le cas.

Les coefficients non paramétriques atteignent les bornes  $\pm 1$  dès qu'il existe une dépendance fonctionnelle entre les variables. Toutefois, le coefficient de corrélation de Pearson est plus restrictif dans sa mesure de la dépendance, en ce qu'il n'égale  $\pm 1$  que si les variables sont une fonction linéaire l'une de l'autre. Un autre avantage des coefficients non paramétriques est qu'ils sont théoriquement toujours bien définis, ce qui n'est pas le cas pour le coefficient de corrélation de Pearson.

	<b>Pointe–Volume</b>	<b>Pointe–Durée</b>	<b>Volume–Durée</b>
Corrélation de Pearson	0,6601	0,0353	0,6574
Rho de Spearman	0,5930	0,0384	0,6569
Tau de Kendall	0,4301	0,0123	0,4940

TAB. 3.1 – Mesures de la dépendance entre la pointe, le volume et la durée de l'Ashuapmushuan

On observe dans le tableau 3.1 que les trois statistiques sont cohérentes dans leur estimation du degré de dépendance entre les variables mesurées sur l'Ashuapmushuan. Il est à noter que sous l'hypothèse nulle d'indépendance entre chacune des paires de

variables, le rho de Spearman suit une loi normale de moyenne 0 et de variance  $1/(n-1)$ , où  $n$  est la taille de l'échantillon. Ceci nous permet de calculer le seuil observé associé à cette statistique. De même, le tau de Kendall suit une loi normale de moyenne 0 et de variance  $2(2n+5)/\{9n(n-1)\}$ . On obtient donc des seuils observés qui indiquent une forte dépendance entre les variables pointe-volume et volume-durée. Cependant, cette dépendance est non significative dans le cas des variables pointe-durée.

Les relations entre les variables pointe, volume et durée sont illustrées deux à deux à la figure 3.1. Les trois graphiques de la première ligne montrent des nuages de points des rangs des variables à partir desquels on peut avoir une idée a priori de la valeur du rho de Spearman ou du tau de Kendall.

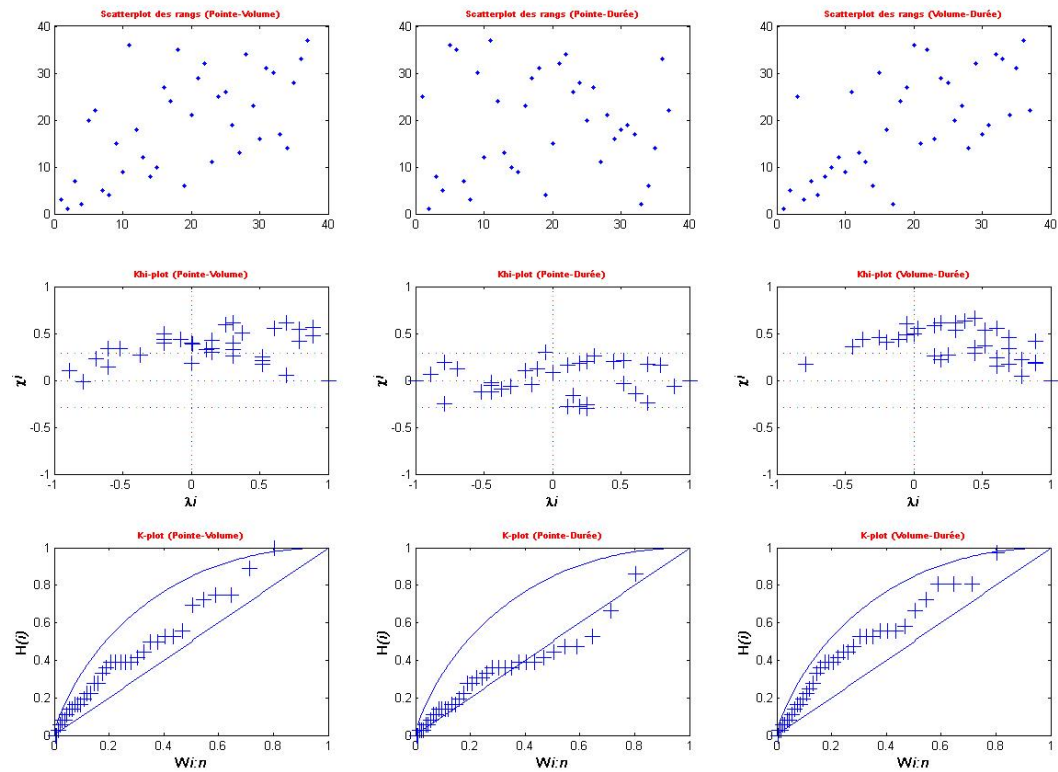


FIG. 3.1 – Graphiques de nuages de points, khi-plot et K-plot illustrant la dépendance entre la pointe, le volume et la durée pour l'Ashuapmushuan

Pour mieux appréhender la dépendance, on peut également construire un khi-plot. Ce graphique, introduit par Fisher et Switzer (1985), est inspiré des cartes de contrôle. Pour construire un khi-plot, on commence d'abord par calculer, pour chaque couple  $(X_i, Y_i)$ , les trois statistiques suivantes :

$$\begin{aligned} H_i &= \#\{j \neq i : X_j \leq X_i, Y_j \leq Y_i\}/(n-1), \\ F_i &= \#\{j \neq i : X_j \leq X_i\}/(n-1), \\ G_i &= \#\{j \neq i : Y_j \leq Y_i\}/(n-1). \end{aligned}$$

Sous l'hypothèse d'indépendance, on s'attendrait à avoir  $H_i \approx F_i \times G_i$ . C'est cette caractéristique que le khi-plot cherche à exploiter en représentant sur un graphique les paires de points  $(\lambda_i, \chi_i)$ , où

$$\chi_i = \frac{H_i - F_i G_i}{\sqrt{F_i(1-F_i)G_i(1-G_i)}}, \quad (3.4)$$

et

$$\lambda_i = 4 \operatorname{signe}(\tilde{F}_i \tilde{G}_i) \max(\tilde{F}_i^2, \tilde{G}_i^2), \quad (3.5)$$

où  $\tilde{F}_i = F_i - 1/2$  et  $\tilde{G}_i = G_i - 1/2$  pour  $1 \leq i \leq n$ .

L'équation (3.4) donne une mesure de la distance entre le couple  $(X_i, Y_i)$  et le centre du nuage de points. Elle est comprise entre  $-1$  et  $1$ . Il faut aussi remarquer que les paires telles que

$$|\lambda_i| < 4 \left( \frac{1}{n-1} - \frac{1}{2} \right)^2$$

sont généralement enlevées dans le but d'éviter les valeurs aberrantes.

L'équation (3.5) présente la corrélation associée aux  $n-1$  paires  $(X_{ij}, Y_{ij})$  d'indicateurs obtenues en fixant  $X_i$  et  $Y_i$  et en posant

$$X_{ij} = \begin{cases} 1 & \text{si } X_j \leq X_i, \\ 0 & \text{sinon,} \end{cases}$$

et

$$Y_{ij} = \begin{cases} 1 & \text{si } Y_j \leq Y_i, \\ 0 & \text{sinon,} \end{cases}$$

pour tout  $j \neq i$ . Les valeurs de l'équation (3.5) sont également comprises entre  $-1$  et  $1$ .

De plus, une propriété intéressante de la statistique  $\chi_i$  est son étroite liaison avec la statistique du khi-deux. En effet, en multipliant l'équation (3.4) par  $\sqrt{n}$  et en élevant au carré, on obtient la statistique du khi-deux, qui sert généralement à tester l'indépendance dans des tableaux de fréquences.

Sous le postulat d'indépendance, la très grande majorité des points devrait se situer à l'intérieur d'une bande de confiance autour de l'axe horizontal. Si toutefois la dépendance est forte, la statistique  $\chi_i$  prendra de grandes valeurs positives ou négatives. De fait, les bornes indiquées sur les graphiques représentent des limites de contrôle à l'intérieur desquelles devraient se situer environ 95 % des observations sous l'hypothèse d'indépendance.

Les khi-plots de l'Ashuapmushuan sont présentés à la deuxième ligne de la figure 3.1. On peut remarquer que la dépendance est fortement positive pour l'interaction pointe et volume ainsi que pour l'interaction volume et durée. Quant à l'interaction pointe et durée, elle présente une très faible dépendance.

Le K-plot, proposé récemment par [Genest et Boies \(2003\)](#), est une autre méthode graphique permettant de visualiser la dépendance. Cette technique s'inspire plutôt de la droite de Henry. On peut construire le K-plot en trois temps, en procédant comme suit :

1. Calculer  $H_i = \#\{j \neq i : X_j \leq X_i, Y_j \leq Y_i\} / (n - 1)$ .
2. Ordonner les  $H_i$  tel que  $H_{(1)} \leq \dots \leq H_{(n)}$ .
3. Tracer les paires  $(W_{i:n}, H_{(i)})$ ,  $1 \leq i \leq n$ , où  $W_{i:n}$  représente l'espérance de la  $i$ ème statistique d'ordre d'un échantillon de taille  $n$  de loi  $K_0(w) = w - w \log(w)$  pour  $w \in [0, 1]$ . [Genest et Boies \(2003\)](#) montrent que  $K_0$  est la loi asymptotique de  $H_i$  sous l'hypothèse d'indépendance.

Le K-plot est une méthode de visualisation non paramétrique de la dépendance, puisqu'elle est basée sur les rangs des données. Donc, plus les données sont près de la diagonale, plus la dépendance est faible. Au contraire, si les données sont près de la borne supérieure, la dépendance est forte. Une dépendance négative extrême de  $-1$  serait illustrée par une droite horizontale égale à 0.

Comme on pourrait s'y attendre, les K-plots de l'Ashuapmushuan sont cohérents avec les khi-plots et les mesures statistiques de dépendance. Tous indiquent une dépendance positive entre la pointe et le volume, ainsi que le volume et la durée, et un faible degré de dépendance entre la pointe et la durée.

## 3.2 Synthèse

Une analyse de la dépendance a également été effectuée pour les trois autres bassins versants. Les mesures d'association résultantes sont présentées aux tableaux 3.2, 3.3 et

3.4. On remarque particulièrement les valeurs obtenues pour la Batiscan et la Romaine, qui sont fortement négatives en ce qui a trait à l'interaction pointe et durée, mais positives pour les interactions pointe et volume ainsi que volume et durée. Ces deux cas justifient l'emploi d'une analyse trivariée, qui pourrait tenir compte de cette forte dépendance.

Rivière	Pointe–Volume	Pointe–Durée	Volume–Durée
Ashuapmushuan	0,6601	0,0353	0,6574
Batiscan	0,4463	−0,3191	0,2882
Harricana	0,7859	−0,1417	0,3474
Romaine	0,5832	−0,4756	0,2124

TAB. 3.2 – Coefficient de corrélation de Pearson entre la pointe, le volume et la durée pour les quatre bassins versants

Rivière	Pointe–Volume	Pointe–Durée	Volume–Durée
Ashuapmushuan	0,5930	0,0384	0,6569
Batiscan	0,5306	−0,2886	0,2145
Harricana	0,6961	−0,1617	0,3880
Romaine	0,5892	−0,3943	0,2260

TAB. 3.3 – Rho de Spearman entre la pointe, le volume et la durée pour les quatre bassins versants

Rivière	Pointe–Volume	Pointe–Durée	Volume–Durée
Ashuapmushuan	0,4301	0,0123	0,4940
Batiscan	0,3816	−0,1893	0,1576
Harricana	0,5221	−0,0966	0,2622
Romaine	0,4070	−0,2812	0,1471

TAB. 3.4 – Tau de Kendall entre la pointe, le volume et la durée pour les quatre bassins versants

# Chapitre 4

## Aspects théoriques des copules méta-elliptiques

Vu la corrélation entre la pointe, le volume et la durée, il est clair qu'une analyse de chacune de ces variables prises séparément est inadéquate. Une façon naturelle de pallier ce problème consiste à recourir à des lois multidimensionnelles, telles la loi normale ou la loi gamma multivariée. Cependant, les lois tridimensionnelles classiques imposent pour ainsi dire toutes l'emploi de lois marginales de même type pour les trois variables. Or, cette condition est irréaliste dans notre cas, puisque nous avons déjà vu au chapitre 2 que les lois univariées de la pointe, du volume et de la durée de la crue d'un même bassin versant sont très différentes. L'approche par copules, qui a beaucoup gagné en popularité ces dernières années, offre une avenue plus prometteuse, que nous avons choisi d'explorer.

### 4.1 Définition d'une copule

Une copule est une fonction de répartition  $C$  à  $p \geq 2$  dimensions dont chacune des marges est uniforme sur l'intervalle  $(0, 1)$ . Étant donné une telle copule, on peut facilement construire une loi multivariée de marges  $F_1, \dots, F_p$  arbitraires en posant

$$F(x_1, \dots, x_p) = C\{F_1(x_1), \dots, F_p(x_p)\}, \quad x_1, \dots, x_p \in \mathbb{R}. \quad (4.1)$$

Si par exemple  $X_1, X_2, X_3$  représentent respectivement la pointe, le volume et la durée d'une crue, il est alors possible, grâce à cette approche, de construire une loi ayant pour marges les trois distributions univariées identifiées au chapitre 2.

De plus, un résultat obtenu par Sklar (1959) montre que cette façon de procéder est en fait complètement générale, en ce sens que toute fonction de répartition  $F$  à  $p$  dimensions peut s'écrire sous la forme (4.1). De plus, la copule en question est *unique* dans la mesure où les marges  $F_1, \dots, F_p$  sont continues.

**Théorème 1** (Sklar 1959). *Soit  $F$  une loi de probabilité de dimension  $p$  ayant pour fonctions de répartition marginales continues  $F_1, \dots, F_p$ . Alors  $F$  possède une représentation unique sous forme de copule, à savoir :*

$$F(x_1, \dots, x_p) = C\{F_1(x_1), \dots, F_p(x_p)\}, \quad x_1, \dots, x_p \in \mathbb{R}.$$

Les livres de Joe (1997) et de Nelsen (1999) présentent une introduction générale aux copules. De plus, on peut référer au livre de Cherubini *et al.* (2004) et à l'article de référence de Frees et Valdez (1998) pour des applications au domaine de la finance et de l'actuariat. Dans le domaine spécifique de l'hydrologie, une première introduction à la modélisation par copules a été présentée par Favre *et al.* (2004) ; un exemple détaillé d'inférence non paramétrique pour des modèles de copules est également présenté dans l'article de Genest et Favre (2006).

### 4.1.1 Copules archimédiennes

Il existe plusieurs familles de copules utilisées pour modéliser un ensemble de variables. La classe des copules archimédiennes, popularisée par Genest et MacKay (1986), est celle qui est le plus souvent utilisée en pratique, compte tenu de ses propriétés analytiques intéressantes et de la facilité avec laquelle on peut les simuler.

On dit qu'une copule  $C$  à  $p$  dimensions est archimédienne s'il existe une fonction continue  $\varphi : [0, 1] \rightarrow [0, \infty]$  strictement décroissante satisfaisant la condition  $\varphi(1) = 0$  telle que

$$C_p(u_1, \dots, u_p) = \varphi^{[-1]}\{\varphi(u_1) + \dots + \varphi(u_p)\}, \quad (4.2)$$

où  $\varphi^{-1}$  représente l'inverse du générateur  $\varphi$ .

Les exemples les plus communs de copules archimédiennes sont les familles de Gumbel, de Frank et de Clayton. Le modèle de Clayton est fréquemment employé en analyse de durées de vie, sous le nom de modèle de fragilité gamma ; voir par exemple Oakes (1989).

Le théorème de Kimberling (1974), énoncé ci-dessous, donne une condition nécessaire et suffisante sur une fonction  $\varphi$  pour qu'elle engendre une copule archimédienne pour toute dimension  $p \geq 2$ .

**Théorème 2** (Kimberling 1974). *Soit  $\varphi : [0, 1] \rightarrow [0, \infty]$  une fonction continue strictement décroissante telle que  $\varphi(0) = \infty$  et  $\varphi(1) = 0$ , et soit  $\varphi^{-1}$  l'inverse de  $\varphi$ . La fonction  $C_p$  définie en (4.2) est alors une copule à  $p$  dimensions pour tout  $p \geq 2$  si et seulement si  $\varphi^{-1}$  est complètement monotone sur  $[0, \infty)$ .*

Malgré leur intérêt, les copules archimédiennes sont très contraignantes, étant donné qu'un seul générateur permet de décrire la dépendance entre toutes les paires de variables. Or comme on l'a vu au chapitre 3, les variables pointe, volume et durée ont des niveaux de dépendance complètement différents. Dans notre cas, il est donc impossible de modéliser conjointement les trois variables au moyen d'une copule archimédienne.

Une généralisation partielle de la notion de copule archimédienne à trois dimensions a été proposée par Joe (1997) en utilisant deux générateurs distincts :

$$C(u_1, u_2, u_3) = \varphi_1^{-1} [\varphi_1 \circ \varphi_2^{-1} \{\varphi_2(u_1) + \varphi_2(u_2)\} + \varphi_1(u_3)].$$

Cependant, pour que cette équation ait les propriétés d'une loi trivariée à marges uniformes, la copule doit respecter la condition suivante :

$$\frac{d}{dt^k} (-1)^{k-1} \varphi_1 \{\varphi_2^{-1}(t)\} \geq 0, \quad k = 1, \dots, \infty.$$

Cet ensemble de contraintes sur les dérivés de  $\varphi_1 \circ \varphi_2^{-1}$  fait en sorte que les paramètres des générateurs, et par conséquent les corrélations entre les variables, se situent dans un intervalle généralement très restreint. De fait, le rapport de Huard *et al.* (2004) confirme que les conditions d'application des copules archimédiennes à la Joe sont tellement exigeantes qu'elles s'adaptent mal aux données de pointe, volume et durée de crue.

## 4.1.2 Copules méta-elliptiques

La classe des lois elliptiques revêt un intérêt pratique considérable, puisqu'elle est facilement applicable en plusieurs dimensions. Plus précisément, une loi est dite elliptique si elle peut être représentée sous la forme

$$\mathbf{Z}_{p \times 1} = \mu_{p \times 1} + R \mathbf{A}_{p \times p} \mathbf{u}_{p \times 1} \sim EC_p(\mu, \Sigma, g), \quad (4.3)$$

où  $\mu$  est un vecteur de localisation,  $R$  est une variable aléatoire positive,  $\mathbf{A}$  est une matrice  $p \times p$  telle que  $\mathbf{A} \mathbf{A}^\top = \Sigma_{p \times p}$ ,  $\mathbf{u}$  est un vecteur uniformément distribué sur la sphère de dimension  $p$  et  $g$  est une fonction d'échelle.

Dans le cas particulier où  $g(t) \propto e^{-t/2}$ , on trouve la loi normale multivariée classique. Voir le tableau 4.1 ou les articles de Fang *et al.* (2002) et de Abdous *et al.* (2005) pour des exemples supplémentaires de générateurs correspondant, entre autres, aux lois de Student, de Cauchy ou de Pearson type II multivariées.

Cependant, les marges des lois elliptiques sont entièrement déterminées par le générateur  $g$  et, au surplus, sont toutes identiques. Néanmoins, en transformant les variables individuellement pour rendre leurs marges uniformes, on obtient une copule dite méta-elliptique, dans laquelle il nous est alors loisible d'injecter, dans un deuxième temps, les marges désirées.

En d'autres termes, on peut construire un vecteur  $X_{p \times 1} = (X_1, \dots, X_p)$  de marges  $F_1, \dots, F_p$  à partir d'un vecteur  $(Z_1, \dots, Z_p)$  de loi elliptique en posant

$$X_i = F_i^{-1}\{Q_g(Z_i)\}, \quad 1 \leq i \leq p$$

où  $Q_g$  est la marge commune des variables  $Z_1, \dots, Z_p$  et  $F_i^{-1}$  représente l'inverse de la fonction de répartition de la loi marginale  $F_i$  de  $X_i$ . Cette construction fondée sur une copule méta-elliptique est notée

$$X_{p \times 1} \sim ME_p(\mu, \Sigma, g; F_1, \dots, F_p).$$

Les articles de Fang *et al.* (2002) et de Abdous *et al.* (2005) présentent diverses propriétés des copules méta-elliptiques. Parmi elles, il importe surtout de savoir pour la suite que le tau de Kendall entre deux composantes d'un vecteur de loi elliptique est indépendant de la fonction  $g$  et est lié à la corrélation  $r$  entre les deux variables par la formule  $r = \sin(\tau\pi/2)$ .

## 4.2 Estimation des paramètres d'une copule méta-elliptique

Les paramètres d'une copule méta-elliptique peuvent être estimés comme suit :

$\mu$  : vecteur de localisation que nous avons considéré, sans restriction de généralité, comme égal à zéro ;

$\Sigma$  : matrice de corrélation

$$\Sigma = \begin{pmatrix} 1 & r_{12} & r_{13} \\ r_{21} & 1 & r_{23} \\ r_{31} & r_{32} & 1 \end{pmatrix}$$

obtenue en posant  $r_{ij} = \sin(\tau_{ij}\pi/2)$ , où le tau de Kendall  $\tau_{ij}$  associé à la paire de variables  $(X_i, X_j)$  est calculé tel qu'expliqué au chapitre 3.

$g$  : fonction d'échelle de la loi elliptique choisie, laquelle peut dépendre d'un paramètre  $m$ , représentant par exemple le nombre de degrés de liberté dans le cas de la loi de Student multivariée ;

$F_1, \dots, F_p$  : fonctions de répartition univariées, telles que déterminées au chapitre 2 ;

Dans notre cas, le paramètre  $m$  de la fonction  $g$  des lois de Student et de Pearson type II multivariées (voir tableau 4.1) a été sélectionné à l'aide d'un outil graphique. Pour ce faire, on a eu recours à des diagrammes de dispersion bivariée représentant une simulation de taille 10 000 d'une loi de la famille choisie avec un paramètre de la fonction  $g$  donné, juxtaposé avec les valeurs observées pour ce cas particulier.

Par exemple, examinons le cas de l'Ashuapmushuan, pour la loi de Student trivariée, présenté à la figure 4.1. Pour chaque nombre  $m$  de degrés de liberté étudié, on a examiné trois graphiques représentant les variables pointe, volume et durée prises deux à deux. Dans chaque cas, les valeurs observées des deux variables sont identifiées par des points de couleur foncée, tandis les valeurs obtenues par simulation sont représentées par des points de couleur pâle. Les valeurs de  $m$  choisies sont celles pour lesquelles nous avons jugé que les simulations répliquaient le mieux possible des données observées, c'est-à-dire  $m = 3$  ou  $m = 4$  dans le cas présent.

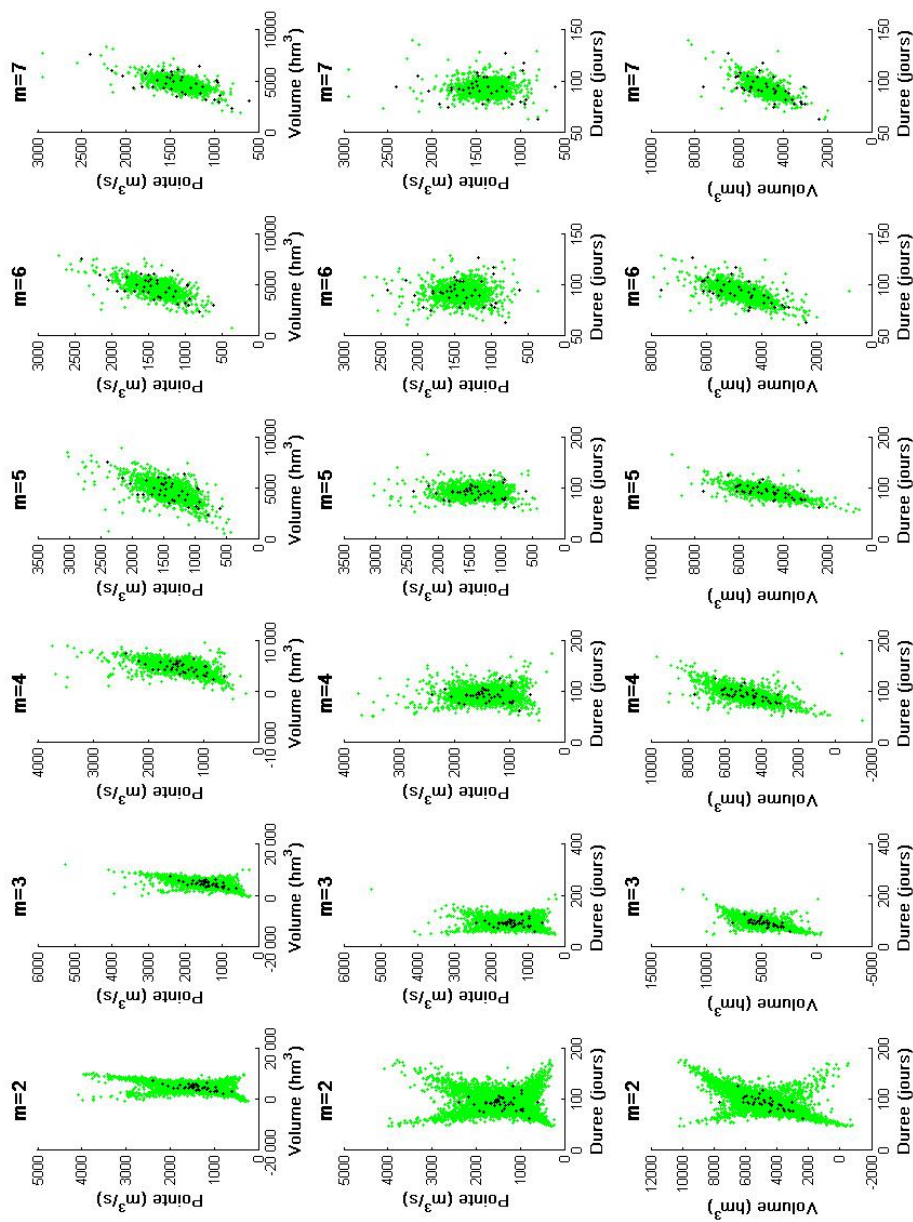


FIG. 4.1 – Représentation graphique de 10 000 simulations d’une loi de marges  $F_1, F_2, F_3$  et de copule méta-elliptique de Student ayant  $m$  degrés de liberté ( $m = 2, 3, 4, 5, 6, 7$ ) superposée aux observations de l’Ashuapmushuan

### 4.3 Algorithme de simulation d'une copule méta-elliptique

La section suivante est basée sur le rapport interne de [Jacques \(2005\)](#) et présente de façon succincte l'algorithme de simulation de copules méta-elliptiques.

Pour simuler un vecteur  $(U_1, \dots, U_p)$  distribué selon une copule méta-elliptique, on doit d'abord fixer la fonction  $g$  de la copule et déterminer les fonctions  $R$  et  $Q_g$ , selon les formules suivantes :

$$f_R(r) = \frac{2\pi^{p/2}}{\Gamma(p/2)} r^{p-1} g(r^2);$$

$$Q_g(x) = \frac{1}{2} + \frac{\pi^{(p-1)/2}}{\Gamma\{(p-1)/2\}} \int_0^x \int_{u^2}^{\infty} (y-u^2)^{(p-1)/2-1} g(y) dy du.$$

Un résumé de ces fonctions pour quatre copules sélectionnées est présenté au tableau [4.1](#).

Ensuite, la simulation peut être effectuée selon l'algorithme suivant, extrait de l'article de [Fang \*et al.\* \(2002\)](#) :

1. Simuler  $\mathbf{u}_{p \times 1}$  uniformément distribué sur la sphère de dimension  $p$ .
2. Générer  $R$ , de loi  $f_R$ , telle que présentée au tableau [4.1](#).
3. Calculer  $\mathbf{A}_{p \times p}$ , la matrice obtenue par la décomposition de Choleski de  $\Sigma$ .
4. Calculer  $\mathbf{Z} = R\mathbf{A}\mathbf{u}$ .
5. Calculer  $U_i = Q_g(Z_i)$ ,  $i = 1, \dots, p$ .
6. Livrer  $(U_1, \dots, U_p)$ .

Si toutefois on désire obtenir un vecteur  $(X_1, \dots, X_p)$  de marges  $F_1, \dots, F_p$ , il suffit alors de poser

$$X_i = F_i^{-1}(U_i), \quad 1 \leq i \leq p.$$

TAB. 4.1 – Fonctions requises pour la simulation de quatre copules méta-elliptiques

Copule	$R^2$	fonction $g$	$Q_g(\mathbf{x})$	Conditions
Normale	$R^2 = \chi^2(n)$	$g(t) = (2\pi)^{-n/2} \exp(-t/2)$	$F_{N(0,1)}(x)$	
Student	Non requis	$g(t) = \frac{(\pi m)^{-n/2} \Gamma(\frac{1}{2}(m+n))}{\Gamma(m/2)} (1 + \frac{t}{m})^{-\frac{1}{2}(m+n)}$	$F_{\text{Student}}(x)$	$N = \frac{1}{2}(m+n)$
Cauchy	Non requis	$g(t) = \frac{(\pi)^{-n/2} \Gamma(\frac{1}{2}(1+n))}{\Gamma(1/2)} (1+t)^{-\frac{1}{2}(1+n)}$	$F_{\text{Cauchy}}(x)$	$N = \frac{1}{2}(1+n),$ $m = 1$
Pearson type II	$R^2 = \text{Beta}(n/2, m+1)$	$g(t) = \frac{\Gamma(n/2+m+1)}{\pi^{n/2} \Gamma(m+1)} (1-t)^m$	$F_{PII}(x)$	$-1 \leq t \leq 1,$ $m > -1$

## 4.4 Exemples d'applications

Les applications utilisant les copules sont de plus en plus fréquentes dans divers domaines de la littérature scientifique. Les copules archimédiennes sont très employées, mais on voit aussi de plus en plus d'articles se servant des copules méta-elliptiques. Voici quelques exemples provenant du domaine de la finance :

- [Lauprete \*et al.\* \(2002\)](#) traitent de l'estimation d'un portefeuille minimisant les risques à l'aide d'une approche robuste. En plus de la copule normale multivariée, ils examinent la modélisation par une copule méta-elliptique de Student.
- [Breymann \*et al.\* \(2003\)](#) analysent la structure de dépendance de données de haute fréquence à deux dimensions en fonction du temps. Une attention particulière est portée au comportement des queues des distributions, lequel est modélisé par des copules de Gumbel, Clayton, Frank, Student et normale.
- [Malevergne et Sornette \(2003\)](#) portent un regard critique sur le postulat à l'effet que la dépendance entre les capitaux financiers est normalement distribuée. Ils illustrent les dangers liés à une utilisation non-éclairée de ce postulat et montrent, en particulier, que la copule méta-elliptique de Student pourrait constituer une alternative intéressante dans certaines situations.
- [Schloegl et O'Kane \(2005\)](#) modélisent la distribution des pertes associées à un portefeuille d'actifs à l'aide des copules de Student, normale, Clayton et Gumbel. Ils comparent les valeurs-à-risques obtenues dans tous les cas.

Il est intéressant de noter que les seules copules méta-elliptiques actuellement employées dans la littérature sont celles liées aux lois normale et de Student multivariées. Nous n'avons pu trouver aucun exemple d'application avec les copules de Pearson type II ou de Cauchy.

# Chapitre 5

## Tests d'adéquation

### 5.1 Tests d'adéquation pour les copules

Les tests d'adéquation pour les copules sont relativement récents. En effet, on retrouve encore peu d'articles sur le sujet, mais le domaine est en constant développement. Tout d'abord, citons [Genest et Rivest \(1993\)](#) qui introduisent le processus de Kendall,

$$\mathbb{K}_n(t) = \sqrt{n}\{K_n(t) - K(t)\}, \quad 0 \leq t \leq 1$$

qu'ils utilisent comme outil de sélection d'une copule archimédienne. Dans l'équation ci-dessus,

$$K(t) = \Pr\{F(X_1, \dots, X_p) \leq t\} = \Pr\{C(U_1, \dots, U_p) \leq t\} = \Pr(W \leq t), \quad 0 \leq t \leq 1$$

est la fonction de répartition de la transformation intégrale de probabilité  $W = F(X_1, \dots, X_p) = C(U_1, \dots, U_p)$  associée au vecteur  $(X_1, \dots, X_p)$  de loi  $F$ , de marges  $F_1, \dots, F_p$  et de copule  $C$ . Quant à  $K_n$ , ce n'est rien d'autre qu'une version expérimentale de la loi de Kendall, déduite d'un échantillon  $(X_{11}, \dots, X_{1p}), \dots, (X_{n1}, \dots, X_{np})$  de loi  $F$ . Plus spécifiquement,

$$K_n(t) = \frac{1}{n} \sum_{i=1}^n \mathbb{I}(W_i \leq t),$$

où pour chaque  $1 \leq i \leq n$ ,

$$W_i = F_n(X_{i1}, \dots, X_{ip})$$

est une pseudo-observation et, de manière générale,

$$F_n(x_1, \dots, x_p) = \frac{1}{n} \sum_{j=1}^n \mathbb{I}(X_{j1} \leq x_1, \dots, X_{jp} \leq x_p).$$

Remarquons que les variables  $W_1, \dots, W_n$  ne sont fonction que des rangs des observations, puisque  $X_{jk} \leq X_{ik}$  se produit si et seulement si le rang de  $X_{jk}$  est inférieur à celui de  $X_{ik}$ .

Le processus de Kendall fut ensuite étudié plus à fond par [Barbe et al. \(1996\)](#) et l'idée de tester l'adéquation d'une copule à l'aide de ce processus fut éventuellement développée par [Wang et Wells \(2000\)](#). Ces derniers en sont venus à établir une procédure de sélection d'une copule archimédienne, pour des données bivariées censurées, en mesurant la distance entre  $K_n(t)$  et  $K(t)$  sous l'hypothèse nulle que  $C$  appartient à une famille paramétrique de copules. Ce premier test d'adéquation fut basé sur une version tronquée du processus de Kendall,

$$S_{\xi n} = \int_{\xi}^1 \{\mathbb{K}_n(t)\}^2 dt, \quad \xi > 0.$$

[Genest et al. \(2006\)](#) ont ensuite élargi ces travaux en montrant comment calculer un seuil observé asymptotique pour deux tests d'adéquation fondés sur une version non tronquée du processus de Kendall, soit

$$S_n = \int_0^1 |\mathbb{K}_n(t)|^2 k(\theta_n, t) dt \quad \text{et} \quad T_n = \sup_{0 \leq t \leq 1} |\mathbb{K}_n(t)|,$$

où  $k(\theta_n, t)$  est la densité associée à  $K_{\theta}(t)$ .

La statistique  $S_n$  est fondée sur la distance de Cramér–von Mises, tandis que la statistique  $T_n$  est basée sur celle de Kolmogorov–Smirnov. Des formules utilisant les rangs des observations permettent d'effectuer ces calculs facilement. De plus, l'application de ces tests n'est ni restreinte aux copules archimédiennes, ni aux problèmes en dimension deux.

En effectuant un bootstrap paramétrique des statistiques  $S_n$  et  $T_n$ , il est possible d'obtenir des seuils approximatifs associés à ces statistiques pour des hypothèses du genre  $C \in (C_{\theta})$ , c'est-à-dire du type : la copule  $C$  associée à  $F$  appartient à la famille de copules  $(C_{\theta})$ . Dans le document de travail de [Genest et Rémillard \(2006\)](#), on confirme la validité de cette approche par bootstrap pour les deux types de tests d'adéquation les plus fréquents, soit ceux où l'on compare la distance entre une distribution multivariée empirique et son estimation paramétrique sous l'hypothèse nulle, ou encore ceux où l'on compare la distance entre les estimations empiriques et paramétriques de pseudo-observations univariées telles que les  $W_i$ , obtenues grâce à la transformation intégrale de probabilité. De plus, pour contrer une situation comme la nôtre où la distribution théorique  $K_{\theta_n}$  n'a pas de forme explicite, [Genest et Rémillard \(2006\)](#) ont expliqué et

validé un test d'adéquation utilisant un bootstrap paramétrique à deux niveaux, lequel est décrit à la section suivante.

Dans le même ordre d'idées, [Fermanian \(2005\)](#) a aussi développé un test d'adéquation pour les copules, basé sur l'estimation du noyau de la fonction de densité de la copule. Ce test suit une distribution asymptotiquement khi-carré, mais son emploi nécessite que l'utilisateur fixe arbitrairement certaines quantités, notamment un noyau et une pondération, ce qui n'est pas toujours évident.

[Chen et Fan \(2005\)](#) ont également proposé un test du rapport des pseudo-vraisemblances pour la sélection de modèles de copules multivariées semiparamétriques. Ce test, qui sert surtout à comparer différentes familles de copules entre elles, ne sera pas considéré dans la suite.

### 5.1.1 Algorithme du bootstrap paramétrique à deux niveaux

Pour tester l'adéquation d'un ensemble de données à une certaine famille de copules méta-elliptiques, nous allons utiliser le test de [Genest \*et al.\* \(2006\)](#), inspiré de la statistique de Cramér-von Mises, à savoir

$$S_n = \int_0^1 |\mathbb{K}_n(t)|^2 k_{\theta_n}(t) dt.$$

Cette statistique est fondée sur une mesure de distance pondérée entre la distribution empirique des pseudo-observations  $W_i$  et une estimation de la distribution  $K_\theta$  associée à la copule méta-elliptique la plus plausible dans la famille postulée sous  $H_0$ . Dans le but d'obtenir un seuil asymptotique associé à cette statistique, nous allons utiliser le bootstrap paramétrique à deux niveaux, tel que décrit dans le document de travail de [Genest et Rémillard \(2006\)](#), pour le cas où la distribution théorique n'a pas de forme explicite.

Tout d'abord, nous disposons d'un échantillon de  $n$  observations pour chacune des trois variables : pointe, volume et durée, respectivement dénotées  $X_1, X_2$  et  $X_3$ . Chaque variable est distribuée selon une loi marginale  $F_p$ , où  $p = 1, 2$  ou  $3$ , telle que déterminée au chapitre 2. Une fois muni de cette notation, l'algorithme du bootstrap paramétrique à deux niveaux se formule ainsi :

1. Construire un vecteur aléatoire uniforme  $\mathbf{U}_i$ ,  $i \in \{1, \dots, n\}$ , en appliquant les fonctions de répartition respectives à chacune des variables  $X_p$

$$\mathbf{U}_i = (U_{1i}, U_{2i}, U_{3i}) = (F_1(X_{1i}), F_2(X_{2i}), F_3(X_{3i})).$$

2. Calculer le tau de Kendall  $\tau_{ij}$  entre chacune des trois paires  $(i, j)$  de variables pointe, volume et durée. En déduire la matrice de corrélation suivante :

$$\Sigma_n = \begin{pmatrix} 1 & r_{12} & r_{13} \\ r_{21} & 1 & r_{23} \\ r_{31} & r_{32} & 1 \end{pmatrix}$$

où  $r_{ij} = \sin(\pi\tau_{ij}/2)$  pour  $i \neq j$ . Cette matrice symétrique, estimée à partir des rangs, joue le rôle du paramètre  $\theta$  générique évoqué dans la section précédente. Dans la suite, on écrit donc indifféremment  $\Sigma_n = \theta_n$ .

3. Calculer la valeur de la statistique  $S_n$ , à l'aide du jeu de données initial,

$$\begin{aligned} S_n &= \frac{n}{3} + n \sum_{j=1}^{n-1} K_n^2 \left( \frac{j}{n} \right) \left\{ K_{\theta_n} \left( \frac{j+1}{n} \right) - K_{\theta_n} \left( \frac{j}{n} \right) \right\} \\ &\quad - n \sum_{j=1}^{n-1} K_n \left( \frac{j}{n} \right) \left\{ K_{\theta_n}^2 \left( \frac{j+1}{n} \right) - K_{\theta_n}^2 \left( \frac{j}{n} \right) \right\}. \end{aligned}$$

Pour ce faire :

- (a) Estimer la loi de Kendall de façon non paramétrique en calculant

$$K_n(t) = \frac{1}{n} \sum_{i=1}^n \mathbb{I} \{ W_i = C_n(\mathbf{U}_i) \leq t \},$$

où

$$C_n(u_1, u_2, u_3) = \frac{1}{n} \sum_{i=1}^n \mathbb{I} \{ U_{1i} \leq D_{1n}^{-1}(u_1), U_{2i} \leq D_{2n}^{-1}(u_2), U_{3i} \leq D_{3n}^{-1}(u_3) \}$$

et

$$D_{jn}(u_j) = \frac{1}{n} \sum_{i=1}^n \mathbb{I} (U_{ji} \leq u_j).$$

- (b) Estimer la loi de Kendall de façon non paramétrique sous l'hypothèse que  $C$  appartient à une famille paramétrique  $(C_\theta)$  de copules (ce qui entraîne que  $K \in (K_\theta)$ ) :

$$K_{\theta_n}(t) = \frac{1}{n} \sum_{i=1}^n \mathbb{I} \{ C_{\theta_n}(\mathbf{U}_i) \leq t \},$$

où  $C_{\theta_n}(u_1, u_2, u_3)$  est donné par

$$\frac{1}{\sqrt{\det(\Sigma_n)}} \int_{-\infty}^{Q_g^{-1}(u_1)} \int_{-\infty}^{Q_g^{-1}(u_2)} \int_{-\infty}^{Q_g^{-1}(u_3)} g(z^\top \Sigma_n^{-1} z) dz_1 dz_2 dz_3.$$

4. Répéter 10 000 fois :

- (a) Générer  $\mathbf{U}_1^*, \dots, \mathbf{U}_n^* \sim C_{\theta_n^*}$ , une simulation de  $n$  observations d'une copule méta-elliptique d'ordre 3, selon l'algorithme présenté au chapitre 4.
- (b) Estimer la matrice de corrélation  $\Sigma_n^* = \theta_n^*$  de  $\mathbf{U}^*$ , telle que décrite au point 2.
- (c) Générer  $\mathbf{U}_1^{**}, \dots, \mathbf{U}_n^{**} \sim C_{\theta_n^{**}}$ , une deuxième simulation d'une copule méta-elliptique d'ordre 3, selon les paramètres de  $\mathbf{U}^*$ , évalués au point précédent.
- (d) Estimer la matrice de corrélation  $\Sigma_n^{**} = \theta_n^{**}$  de  $\mathbf{U}^{**}$ , selon la procédure décrite au point 2.
- (e) Calculer la valeur de la statistique  $S_n^*$  pour chaque simulation effectuée,

$$S_n^* = \frac{n}{3} + n \sum_{j=1}^{n-1} K_n^{*2} \left( \frac{j}{n} \right) \left\{ K_{\theta_n^{**}} \left( \frac{j+1}{n} \right) - K_{\theta_n^{**}} \left( \frac{j}{n} \right) \right\} \\ - n \sum_{j=1}^{n-1} K_n^* \left( \frac{j}{n} \right) \left\{ K_{\theta_n^{**}} \left( \frac{j+1}{n} \right) - K_{\theta_n^{**}} \left( \frac{j}{n} \right) \right\}$$

Pour ce faire :

i. Calculer

$$K_n^*(t) = \frac{1}{n} \sum_{i=1}^n \mathbb{I} \{ C_n^*(\mathbf{U}_i^*) \leq t \}$$

où

$$C_n^*(u) = \frac{1}{n} \sum_{i=1}^n \mathbb{I} \{ U_{1i}^* \leq D_{1n}^{*-1}(u_1), U_{2i}^* \leq D_{2n}^{*-1}(u_2), U_{3i}^* \leq D_{3n}^{*-1}(u_3) \}$$

et

$$D_{jn}^*(u_j) = \frac{1}{n} \sum_{i=1}^n \mathbb{I} (U_{ji}^* \leq u_j)$$

ii. Calculer

$$K_{\theta_n^{**}}(t) = \frac{1}{n} \sum_{i=1}^n \mathbb{I} \{ C_{\theta_n^{**}}(\mathbf{U}_i^{**}) \leq t \},$$

où  $C_{\theta_n^{**}}(u_1, u_2, u_3)$  est donné par

$$\frac{1}{\sqrt{\det(\Sigma_n^{**})}} \int_{-\infty}^{Q_g^{-1}(u_1)} \int_{-\infty}^{Q_g^{-1}(u_2)} \int_{-\infty}^{Q_g^{-1}(u_3)} g(z^\top \Sigma_n^{**^{-1}} z) dz_1 dz_2 dz_3.$$

5. Estimer la valeur critique. Pour ce faire, ordonner les statistiques  $S_n^*$  :

$$S_{1:N}^* \leq \dots \leq S_{N:N}^*.$$

La valeur critique de seuil  $\alpha$  sera égale à  $S_{[(1-\alpha)N]:N}^*$ .

6. Estimer le seuil par

$$\text{Seuil} = \frac{1}{N} \# \{j : S_j^* \leq S_n\}$$

où la valeur de  $S_n$  a été calculée au point 3.

## 5.2 Implantation

L'implantation informatique de l'algorithme du bootstrap paramétrique à deux niveaux a été réalisée à l'aide du logiciel Matlab. En cours d'implantation, nous avons éprouvé quelques difficultés à évaluer les intégrales triples intervenant dans le calcul des statistiques  $S_n$  et  $S_n^*$ . En particulier, le temps d'exécution de la fonction `triplequad` était extrêmement long. Dans le cas de la Batiscan, par exemple, le calcul d'une seule intégrale pouvait souvent dépasser une minute. Étant donné que l'on voulait effectuer 10 000 simulations pour les quatre types de copules avec plusieurs choix de paramètres pour chacune d'elles, ça revenait à calculer plus de 70 000 intégrales triples pour chaque rivière, ce qui aurait pris un temps énorme.

Par ailleurs, puisque le domaine d'intégration était non borné, le problème se posait de tronquer ces intégrales sans pour autant affecter la précision du calcul. Plus les bornes inférieures choisies étaient petites, plus le temps d'intégration augmentait, mais plus l'intégrale était précise, d'où le dilemme. Nous avons donc cherché à diminuer le temps d'exécution des intégrales triples. Pour ce faire, nous avons d'abord opté pour le changement de variable suivant :

$$z_i = \begin{cases} a_i \left( 2 - \frac{1}{y_i} \right), & i = 1, 2, 3, \quad \text{si } a_i > 0; \\ a_i \left( \frac{1}{y_i} \right), & i = 1, 2, 3, \quad \text{si } a_i < 0. \end{cases} \quad (5.1)$$

Ceci a eu pour effet de transformer le domaine d'intégration en un intervalle borné, à savoir  $[0, 1]$ . Par la suite, nous avons utilisé un code en C pour effectuer l'intégration triple de la copule normale et de celle de Pearson type II, ce qui a grandement diminué le temps de calcul. Celui-ci s'est alors avéré être de l'ordre de grandeur d'à peine une seconde. Malheureusement, cette fonction ne permettait pas de traiter correctement des cas plus complexes comme les copules de Student et de Cauchy. Pour ces deux cas, nous avons donc plutôt réduit le nombre de simulations à 5 000 et réalisé des analyses de sensibilité dans ces situations.

# Chapitre 6

## Application

### 6.1 Résultats

Le test d'adéquation pour les copules méta-elliptiques, tel que présenté au chapitre précédent, a été appliqué aux données de chacune des quatre rivières à l'étude. Dans chaque cas, quatre familles de copules ont été ajustées, en utilisant différents paramètres de la fonction  $g$ . Ces paramètres ont été choisis à l'aide d'une méthode graphique, telle que décrite au chapitre 4.

Dans le cas de la copule de Pearson type II, trois valeurs possibles du paramètre de la fonction  $g$  ont été considérées pour chaque rivière, soit 2, 3 et 4 degrés de liberté. Le temps d'exécution pour la copule de Student étant beaucoup plus grand, on s'est contenté d'essayer deux de ces trois valeurs. Pour leur part, les copules de Cauchy et normale ne possèdent aucun paramètre.

Au départ, le nombre de simulations souhaitées pour le bootstrap avait été fixé à 10 000 pour chaque ajustement. Cependant, le temps d'exécution très grand nous a obligés à réduire ce nombre à 5 000 dans le cas des deux copules qui n'avaient pu être programmées en C, soit la copule de Student et celle de Cauchy. Une analyse de sensibilité du seuil observé nous a permis de déduire que le nombre de simulations réalisées était tout de même assez grand, puisque le seuil observé convergeait toujours après peu d'itérations.

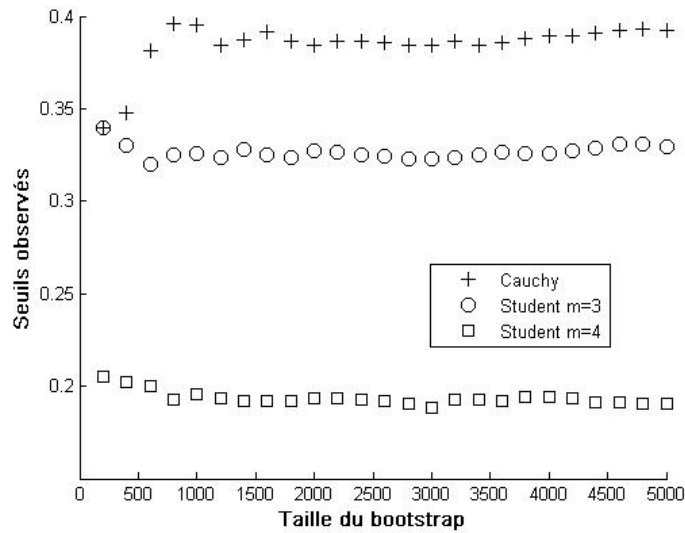


FIG. 6.1 – Analyse de sensibilité des ajustements des copules de Cauchy et Student, avec paramètre  $m = 3$  et 4, selon les données de l’Ashuapmushuan

Le graphique 6.1 présente les résultats de l’analyse de sensibilité effectuée pour les données de l’Ashuapmushuan. On observe que même en diminuant le nombre de simulations, le seuil observé varie très peu. Pour le cas présenté, ce seuil varie d’un maximum de 1% selon que le nombre de simulations soit 1000 ou 5000.

Le tableau 6.1 présente les résultats de toutes les simulations exécutées. Si le seuil du test d’adéquation est supérieur au seuil fixé, c’est-à-dire 0.05, on ne rejette pas l’hypothèse nulle et on conclut que la copule avec ce paramètre s’ajuste bien aux données. En revanche, si le seuil est inférieur à 0.05, on rejette l’hypothèse nulle et on conclut que la copule avec ce paramètre ne convient pas aux données.

Les copules rejetées ont été indiquées en italique dans le tableau 6.1. Étant donné que plusieurs copules semblaient bien s’ajuster aux données, nous avons arrêté notre choix sur la copule présentant le seuil le plus grand, que nous avons indiqué en gras dans le tableau 6.1.

Les seuils observés des rivières Harricana et Ashuapmushuan nous indiquent que la copule de Pearson type II, de paramètre égal à 2, serait la plus appropriée parmi celles étudiées. Pour ce qui est des rivières Batiscan et Romaine, on choisit plutôt la copule de Student, de paramètre égal à 2.

Une représentation graphique de l'ajustement de la copule de Pearson type II de paramètre égal à 2 est présenté à la figure 6.2 pour l'Ashuapmushuan . On constate que les observations simulées par la copule semblent bien reproduire les données originales.

Rivière	Copule	Paramètre	Nb simulations	Seuil observé
Ashuapmushuan	Normale	-	10 000	0,5809
<b>Ashuapmushuan</b>	<b>P2</b>	<b>2</b>	<b>10 000</b>	<b>0,7037</b>
Ashuapmushuan	P2	4	10 000	0,6964
Ashuapmushuan	P2	6	10 000	0,7028
Ashuapmushuan	Student	3	5 000	0,3292
Ashuapmushuan	Student	4	5 000	0,1906
Ashuapmushuan	Cauchy	-	5 000	0,3926
<i>Batiscan</i>	<i>Normale</i>	-	<i>10 000</i>	<i>0,0473</i>
<i>Batiscan</i>	<i>P2</i>	<i>2</i>	<i>10 000</i>	<i>0,0354</i>
Batiscan	P2	4	10 000	0,0684
Batiscan	P2	6	10 000	0,0654
<b>Batiscan</b>	<b>Student</b>	<b>2</b>	<b>5 000</b>	<b>0,0996</b>
<i>Batiscan</i>	<i>Student</i>	<i>3</i>	<i>5 000</i>	<i>0,0528</i>
Batiscan	Cauchy	-	5 000	0,0750
Harricana	Normale	-	10 000	0,5059
<b>Harricana</b>	<b>P2</b>	<b>2</b>	<b>10 000</b>	<b>0,8788</b>
Harricana	P2	4	10 000	0,8738
Harricana	P2	6	10 000	0,8256
<i>Harricana</i>	<i>Student</i>	<i>3</i>	<i>5 000</i>	<i>0,0248</i>
<i>Harricana</i>	<i>Student</i>	<i>4</i>	<i>5 000</i>	<i>0,0072</i>
Harricana	Cauchy	-	5 000	0,0840
Romaine	Normale	-	10 000	0,1666
Romaine	P2	2	10 000	0,1091
Romaine	P2	4	10 000	0,1253
Romaine	P2	6	10 000	0,1261
<b>Romaine</b>	<b>Student</b>	<b>2</b>	<b>5 000</b>	<b>0,1978</b>
Romaine	Student	3	5 000	0,1514
<i>Romaine</i>	<i>Student</i>	<i>4</i>	<i>5 000</i>	<i>0,0450</i>
Romaine	Cauchy	-	5 000	0,1050

TAB. 6.1 – Résultats des ajustements aux copules normale, de Pearson type II, de Student et de Cauchy pour les quatre bassins versants ; les copules choisies sont en gras et les copules rejetées au seuil de 5% sont en italique

Les praticiens sont intéressés par la détermination de quantités statistiques telles que les probabilités conjointes (trivariées), les probabilités conditionnelles ainsi que les temps de retour conjoints. Après avoir choisi une famille de copules, estimé ses paramètres et testé son adéquation, il est possible de calculer ces quantités simplement, directement à partir de l'expression de la copule. [Yue et Rasmussen \(2002\)](#) et [Salvadori et De Michele \(2004\)](#) montrent comment calculer ces probabilités dans le cas bivarié. La généralisation en dimension trois est immédiate.

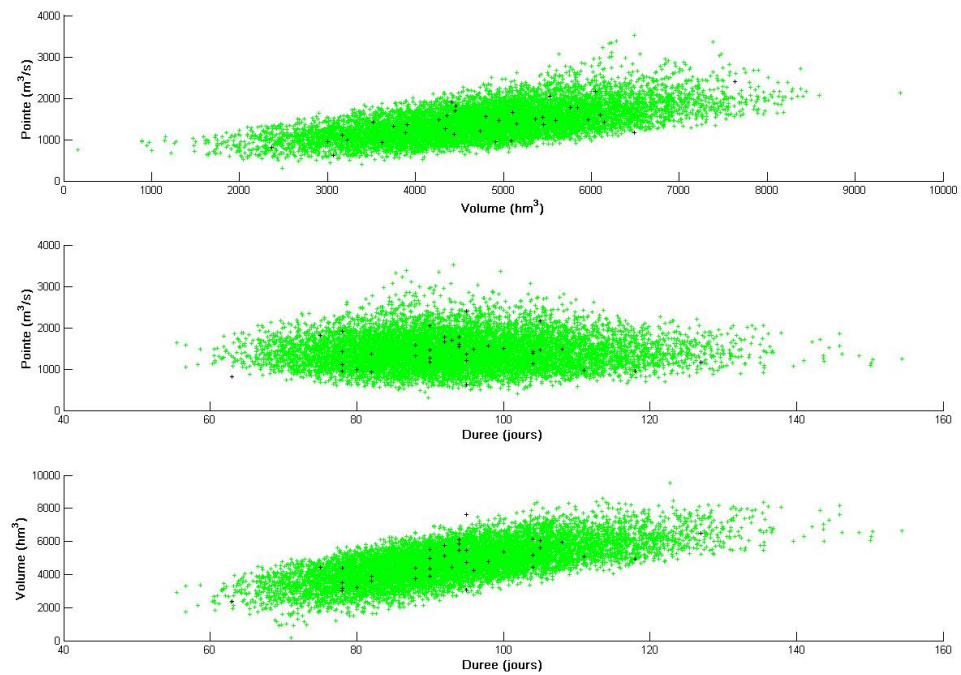


FIG. 6.2 – Représentation graphique de l’ajustement de 10 000 observations de la copule Pearson type II avec paramètre  $m = 2$  superposée aux données originales de l’Ashuapmushuan

# Chapitre 7

## Conclusion

Ce mémoire présente la modélisation multivariée des trois caractéristiques principales des crues, soit la pointe, le volume et la durée. À la suite d'une analyse univariée et d'une analyse de dépendance, diverses familles de copules méta-elliptiques ont été étudiées, soit les copules de Cauchy, normale, Pearson type II et Student. Cette classe a été utilisée en raison de sa flexibilité et du fait qu'elle peut être employée en dimension  $p$  quelconque. Une telle approche permet de mieux estimer le risque lié à un événement hydrologique particulier. En effet, une simple analyse fréquentielle univariée conduit à une sous-estimation ou une surestimation du risque.

L'essentiel de ce travail de maîtrise a consisté à mettre en œuvre une classe de copules encore peu utilisée, ainsi qu'un test d'adéquation très récent. Celui-ci a ensuite été appliqué aux données de quatre bassins versants québécois, soit ceux des rivières Ashuapmushuan, Batiscan, Harricana et Romaine. Plus spécifiquement, un test d'adéquation proposé par [Genest \*et al.\* \(2006\)](#) a été utilisé pour comparer les différents ajustements. Ce test compare la distance de Cramér–von Mises entre la distribution théorique d'intérêt et la distribution échantillonnale de la copule méta-elliptique. Dans notre cas, la distribution théorique ne possédait pas de forme explicite. Cependant, l'utilisation d'un algorithme de bootstrap paramétrique à deux niveaux proposé par [Genest et Rémillard \(2006\)](#) nous a permis de contourner cette difficulté. Ce test nous a également permis d'obtenir des seuils observés et ainsi de pouvoir juger plus facilement de la copule la plus appropriée.

Au vu des résultats de l'analyse fréquentielle présentée, plusieurs familles de copules semblaient adéquates. En dernière analyse, nous avons opté pour la copule dont le seuil observé était le plus grand, soit la copule de Pearson type II de paramètre 2 pour les rivières Harricana et Ashuapmushuan et Student de paramètre 2 pour les rivières

Batiscan et Romaine.

Plusieurs autres familles de copules méta-elliptiques auraient pu être proposées et comparées pour modéliser les variables de pointe, volume et durée de la crue, notamment les copules de Kotz, logistique ou encore la copule de Student généralisée. Toutefois, leurs expressions mathématiques étaient plus complexes et auraient nécessité un développement théorique plus approfondi.

Il faut également souligner que le développement d'algorithmes performants pour des intégrations triples avec singularités ainsi qu'une programmation de toutes les routines d'intégration dans un langage informatique plus performant serait nécessaire pour pouvoir réaliser le test d'adéquation dans un laps de temps raisonnable. Selon nous, le temps de calcul constitue le principal obstacle pour une utilisation opérationnelle de l'approche.

Quoi qu'il en soit, il appert que la classe des copules méta-elliptiques est très flexible. Elle trouvera sans doute plusieurs autres domaines d'application dans le futur.

# Bibliographie

- ABDOUS, B., GENEST, C. et RÉMILLARD, B. (2005). Dependence properties of meta-elliptical distributions. *In Statistical Modeling and Analysis for Complex Data Problems*, publié sous la direction de P. Duchesne et de B. Rémillard, pages 1–15. Springer, New York.
- BARBE, P., GENEST, C., GHOUDI, K. et RÉMILLARD, B. (1996). On Kendall's process. *Journal of Multivariate Analysis*, 58:197–229.
- BREYMAN, W., DIAS, A. et EMBRECHTS, P. (2003). Dependence structures for multivariate high-frequency data in finance. *Quantitative Finance*, 3:1–14.
- CAPÉRAÀ, P. et VAN CUTSEM, B. (1988). *Méthodes et modèles en statistique non paramétrique*. Presses de l'Université Laval et Dunod, Paris.
- CHAIRE EN HYDROLOGIE STATISTIQUE (2002). Hyfran. Rapport technique, INRS-ETE (Université du Québec). Logiciel pour l'analyse fréquentielle en hydrologie, version 1.1.
- CHEN, X. et FAN, Y. (2005). Pseudo-likelihood ratio tests for semiparametric multivariate copula model selection. *La revue canadienne de statistique*, 33:389–414.
- CHERUBINI, U., LUCIANO, E. et VECCHIATO, W. (2004). *Copula Methods in Finance*. Wiley, New York.
- DE MICHELE, C., SALVADORI, G., CANOSSO, M., PETACCIA, A. et ROSSO, R. (2005). Bivariate statistical approach to check adequacy of dam spillway. *Journal of Hydrologic Engineering*, 10:50–57.
- FANG, H.-B., FANG, K.-T. et KOTZ, S. (2002). The meta-elliptical distributions with given marginals. *Journal of Multivariate Analysis*, 82:1–16.
- FAVRE, A.-C., EL ADLOUNI, S., PERREAULT, L., THIÉMONGE, N. et BOBÉE, B. (2004). Multivariate hydrological frequency analysis using copulas. *Water Resources Research*, 40(W01101):doi :10.1029/2003WR002456.

- FERMANIAN, J.-D. (2005). Goodness-of-fit tests for copulas. *Journal of Multivariate Analysis*, 95:119–152.
- FISHER, N. L. et SWITZER, P. (1985). Chi-plots for assessing dependence. *Biometrika*, 72:253–265.
- FREES, E. et VALDEZ, E. A. (1998). Understanding relationships using copulas. *North American Actuarial Journal*, 2:1–25.
- GENEST, C. et BOIES, J.-C. (2003). Detecting dependence with Kendall plots. *The American Statistician*, 57:275–284.
- GENEST, C. et FAVRE, A.-C. (2006). Everything you always wanted to know about copula modeling but were afraid to ask. *Journal of Hydrologic Engineering*, 11:sous presse.
- GENEST, C. et MACKAY, R. J. (1986). Copules archimédiennes et familles de lois bidimensionnelles dont les marges sont données. *La revue canadienne de statistique*, 14:145–159.
- GENEST, C., QUESSY, J.-F. et RÉMILLARD, B. (2006). Goodness-of-fit procedures for copula models based on the probability integral transformation. *Scandinavian Journal of Statistics*, 33:337–366.
- GENEST, C. et RÉMILLARD, B. (2006). Validity of the parametric bootstrap for goodness-of-fit testing in semiparametric models. *Soumis*.
- GENEST, C. et RIVEST, L.-P. (1993). Statistical inference procedures for bivariate Archimedean copulas. *Journal of the American Statistical Association*, 88:1034–1043.
- GOEL, N., SETH, S. et CHANDRA, S. (1998). Multivariate modeling of flood flows. *Journal of Hydraulic Engineering*, 124:146–155.
- HACHÉ, M., PERREAULT, L., RÉMILLARD, L. et BOBÉE, B. (1999). Méthodologie de sélection des distributions statistiques : Application au bassin hydrographique du Saguenay – Lac Saint-Jean. *Revue canadienne de Génie civil*, 26(2):216–225.
- HUARD, D., FAVRE, A.-C., POULIN, A. et ÉVIN, G. (2004). Faisabilité de l'analyse fréquentielle trivariée des pointes, volumes et durées de crue à l'aide des copules archimédiennes. Rapport technique, INRS - Eau, Terre et Environnement, Québec, Canada.
- JACQUES, C. (2004). Analyse univariée des séries de données pour le projet analyse fréquentielle multivariée de la pointe, du volume et de la durée de crue. Rapport technique No I-151, INRS - Eau, Terre et Environnement, Québec, Canada.

- JACQUES, C. (2005). Simulation des copules méta-elliptiques pour le projet analyse fréquentielle multivariée de la pointe, du volume et de la durée de la crue. Rapport technique, INRS - Eau, Terre et Environnement, Québec, Canada.
- JOE, H. (1997). *Multivariate Models and Dependence Concepts*. Monographs on Statistics and Applied Probability. Chapman & Hall, London.
- KIMBERLING, C. (1974). A probabilistic interpretation of complete monotonicity. *Aequationes Mathematicae*, 10:152–164.
- LAUPRETE, G. J., SAMARO, A. M. et WELSCH, R. E. (2002). Robust portfolio optimization. *Metrika*, 55:139–149.
- MALEVERGNE, Y. et SORNETTE, D. (2003). Testing the gaussian copula hypothesis for financial assets dependences. *Quantitative Finance*, 3:231–250.
- NELSEN, R. B. (1999). *An Introduction to Copulas*. Lecture Notes in Statistics. Springer-Verlag, New York.
- OAKES, D. (1989). Bivariate survival models induced by frailties. *Journal of the American Statistical Association*, 84:487–493.
- PRESS, S. J. (2003). *Subjective and Objective Bayesian Statistics*. Wiley, New York, Deuxième édition.
- SALVADORI, G. et DE MICHELE, C. (2004). Frequency analysis via copulas : Theoretical aspects and applications to hydrological events. *Water Resources Research*, 40(W12511):doi :10.1029/2004WR003133.
- SCHLOEGL, L. et O’KANE, D. (2005). A note on the large homogeneous portfolio approximation with the Student-t copula. *Finance and Stochastics*, 9:577–584.
- SINGH, K. et SINGH, V. (1991). Derivation of bivariate probability density functions with exponential marginals. *Stochastic Hydrology and Hydraulics*, 5:55–68.
- SKLAR, A. (1959). Fonctions de répartition à  $n$  dimensions et leurs marges. *Publications de l’Institut de statistique de l’Université de Paris*, 8:229–231.
- WANG, W. et WELLS, M. T. (2000). Model selection and semiparametric inference for bivariate failure-time data. *Journal of the American Statistical Association*, 95:62–76.
- YUE, S. (2000). The bivariate lognormal distribution to model a multivariate flood episode. *Hydrological Processes*, 14:2575–2588.
- YUE, S., OUARDA, T., BOBÉE, B., LEGENDRE, P. et BRUNEAU, P. (1999). The Gumbel mixed model for flood frequency analysis. *Journal of Hydrology*, 226:88–100.

YUE, S. et RASMUSSEN, P. (2002). Bivariate frequency analysis : discussion of some useful concepts in hydrological application. *Hydrological Processes*, 16:2881–2898.