

THESE

En vue de l'obtention du : **DOCTORAT**

Centre de Recherche : CEREMAR

Structure de Recherche : Laboratoire de Mathématiques, Statistique et Applications

Discipline : Mathématiques

Spécialité : Statistique

Présentée et soutenue le 27/6/2022 par :

Mme Wafaa EL HANNOUN

Contribution à la modélisation des variables d'un bassin hydrologique à l'aide des copules et des vine-copules dans un contexte de changement climatique

JURY

Khalid EL HIMDI	PES, Université Mohammed V, Faculté des Sciences - Rabat	Président / Rapporteur
Abdellah OULD KHAL	PES, Université Mohammed V, Faculté des Sciences - Rabat	Rapporteur / Examineur
Abdelouahid IMLAHI	PES, Université Abdelmalek Essaadi, Faculté des Sciences et Techniques - Tanger	Rapporteur / Examineur
Salah-Eddine EL ADLOUNI	Professeur agrégé, Université de Moncton, Faculté des Sciences - Moncton, Canada	Co-directeur de thèse
Abdelhak ZOGLAT	PES, Université Mohammed V, Faculté des Sciences - Rabat	Directeur de thèse

Année Universitaire : 2021/2022

À la mémoire de mes parents
Karima Jirari
&
Omar El Hannoun

DÉDICACE

À la fin de mon parcours de doctorante, que je ne manquerai pas de qualifier de celui d'un combattant, je souhaite dédier mon travail de thèse à un certain nombre de personnes qui m'ont prêté aide, chacun à sa manière, et qui ont cru en ma capacité d'arriver un jour au bout du tunnel.

À mon mari pour sa patience et son support infailible.

À mes enfants Yahya et Hosna pour leur bonne humeur contagieuse.

À mon frère Amine, mon Supporter N°1, mon coach et mon conseiller.

À mon frère Nawfal, mon complice d'enfance et de toujours.

À tous les membres des familles El Hannoun, Jirari et Benzaira pour leurs encouragements et prières.

REMERCIEMENTS

Ce travail de thèse a été réalisé au sein du laboratoire de Mathématiques, Statistique et applications à la Faculté des Sciences de Rabat, Université Mohammed V-Agdal, sous la direction de Monsieur **Abdelhak ZOGLAT**, Professeur de l'enseignement supérieur et responsable dudit laboratoire et sous la co-direction de Monsieur **Salah-Eddine EL ADLOUNI**, Professeur à la Faculté des Sciences, Université de Moncton au Canada.

Tout d'abord, je tiens à adresser ma grande reconnaissance et ma profonde gratitude au Professeur **Abdelhak ZOGLAT** pour le temps énorme qu'il a consacré à mon encadrement, pour ses orientations, son transfert de savoir, ses corrections, son implication et ses encouragements.

J'exprime mes profonds remerciements au Professeur **Salah-Eddine EL ADLOUNI** pour sa contribution majeure dans l'évolution de mes recherches, la pertinence de ses remarques et ses encouragements.

Mes sincères remerciements vont également à Monsieur **Khalid EL HIMDI**, Professeur au sein du Département de Mathématiques à la Faculté des Sciences de Rabat, d'avoir accepté de présider le jury de soutenance et d'avoir bien voulu apporter ses suggestions d'amélioration à mon manuscrit, en sa qualité d'examineur.

Mes remerciements vont aussi à Monsieur **Abdellah OULD KHAL**, Professeur au sein du Département de Mathématiques à la Faculté des Sciences de Rabat, d'avoir accepté de participer au jury de cette thèse en sa qualité d'examineur. Je le remercie aussi pour l'intérêt qu'il a porté à ce travail et sa contribution à l'amélioration de ce manuscrit en sa qualité de rapporteur.

Je remercie aussi Monsieur **Abdelouahid IMLAHI**, Professeur à la Faculté des Sciences et Techniques de Tanger, d'avoir bien voulu participer à l'amélioration de ma thèse en sa qualité de rapporteur et d'examineur.

Je saisis cette occasion pour exprimer ma profonde gratitude à Monsieur Zine-El Abidine Guennoun, Professeur au sein du Département de Mathématiques à la Faculté des Sciences de Rabat et chef dudit Département ainsi qu'au personnel administratif du département pour leur soutien et leurs encouragements.

Je tiens aussi à remercier Madame Fadoua Badaoui, Professeur à l'Institut National de Statistique et d'Economie Appliquée et Monsieur Amine Amar, Professeur au sein de l'Université Al Akhawayn, mes co-auteurs, pour leur contribution effective à mes recherches et leurs précieux conseils.

Je n'oublierai pas d'exprimer ma reconnaissance envers Leila Ait Hassou, pour toute l'aide qu'elle a su m'apporter, en particulier son aide à implémenter mon premier programme en langage *R*.

RÉSUMÉ

Le changement climatique et son impact sur les futures solutions de gestion de l'eau sont devenus cruciaux pour l'avenir de la planète. Les barrages étant une solution primaire d'approvisionnement en eau, différentes approches ont été adoptées pour prédire les effets croissants du changement climatique, en particulier sur l'envasement des barrages. Dans l'absence d'un modèle mathématique fiable de l'envasement qui puisse tenir compte des effets du changement climatique, nous proposons dans cette thèse une approche basée sur les copules pour modéliser et évaluer ces effets sur les variables hydrologiques des réservoirs telles que l'envasement et l'apport en eau. Nos résultats confirment le changement brutal qu'a subi le régime hydrologique de la région de la Moulouya (Maroc) en 1979, illustré par un changement complet de la copule représentant le comportement joint des variables en interaction utilisées. De plus, en utilisant la théorie des copules à valeurs extrêmes, nous mettons en place un modèle pour prévoir l'envasement pour différents scénarios de changement de température. Dans la deuxième partie de cette thèse, nous proposons une méthode basée sur les copules-vignes pour l'évaluation du risque composite. La modélisation de la dépendance basée sur les copules est un outil populaire dans l'évaluation des risques conditionnels, mais elle est généralement appliquée à des paires de variables. En revanche, les copules-vignes offrent une plus grande flexibilité et permettent la modélisation de modèles de dépendance complexes en utilisant la grande variété de copules bivariées qui peuvent être disposées et analysées dans une structure arborescente pour explorer de multiples dépendances. Nous proposons aussi une application de cette approche sur les données de cinq réservoirs situés dans la cascade du bassin de la rivière Saint-Jean dans l'est du Canada. Les résultats des copules-vignes développées conduisent aux périodes de retour jointes et conditionnelles de volumes maximaux, pouvant servir pour la conception hydrologique et la gestion des réservoirs en cascade dans le bassin.

Mots clés : Changement climatique ; copules ; copules-vignes ; envasement ; gestion des barrages ; hydrologie.

ABSTRACT

Climate change and its impact on future water-management-solutions have become crucial to the planet's forthcoming. Since dams are a primary water supply and sourcing solution, different approaches have been taken to predict the increasing climate change effects, especially on dam's siltation. Most of the large-scale dams are periodically controlled through bathymetry measurements. The need for a reliable mathematical model that takes into account climate change effects is still prevalent. In this framework, we use a copula approach to model and evaluate climate change effects on reservoir variables such as siltation and water intake. Our study confirms the abrupt change in the hydrology regime of the Moulouya region (Morocco) back in 1979, by a complete change in the copula representing the joint behaviour of the interacting variables in use. Moreover, using the extreme-value copulas theory we setup a model to forecast siltation for different temperature change scenarios. Besides, this thesis features an application of Regular vine-copulas which are a novel and recently developed statistical and mathematical tool that can be applied in the assessment of composite risk. Copula-based dependence modelling is a popular tool in conditional risk assessment but is usually applied to pairs of variables. By contrast, vine-copulas provide greater flexibility and permit the modelling of complex dependency patterns using the wide variety of bivariate copulas which may be arranged and analysed in a tree structure to explore multiple dependencies. This study emphasises the use of vine-copulas in an analysis of the co-dependencies of five reservoirs in the cascade of the Saint-John River basin in Eastern Canada. The developed R-vine copulas lead to the joint and conditional return periods of maximum volumes, for hydrologic design and cascade reservoir management in the basin. The main attraction of this approach to risk modelling is the flexibility in the choice of distributions used to model heavy-tailed marginals and co-dependencies.

Key words : Climate change ; copulas ; dam management ; hydrology ; siltation ; vine-copulas.

TABLE DES MATIÈRES

Liste des tableaux	i
Table des figures	ii
Acronymes	iv
Introduction générale	1
Chapitre 1 Théorie des copules et copules-vignes	5
1.1 La théorie des copules	6
1.1.1 Théorème de Sklar	6
1.1.2 La copule empirique	7
1.1.3 Familles des copules	8
1.1.3.1 Famille elliptique	8
1.1.3.2 Famille Archimédienne	9
1.1.3.3 Copules de valeurs extrêmes	13
1.1.4 Copules de survie et rotations	15
1.1.5 Copules et formes de dépendance	18
1.1.5.1 La copule d'indépendance	18
1.1.5.2 Relation entre les copules et les mesures de dépendance	18
1.1.5.3 Dépendance des queues de distributions des copules	22
1.2 Construction et estimation des copules	24
1.2.1 Algorithme général	24
1.2.2 Modélisation des marges	24

1.2.3	Les Kendall-plots	26
1.2.4	Test d'indépendance	27
1.2.5	Estimation des paramètres	28
1.2.6	Comparaison des AIC	28
1.2.7	Test d'ajustement	29
1.2.8	Prévision	30
1.3	Les copules-vignes	32
1.3.1	La R-copule-vigne	32
1.3.2	Copules conditionnelles	33
1.3.3	Spécification de la R-copule-vigne	36
1.3.3.1	Algorithme général	36
1.3.3.2	Choix des copules	36
1.3.3.3	Estimation des paramètres	37
1.3.3.4	Comparaison des critères AIC et BIC	38
1.3.3.5	Test d'ajustement de Vuong	38
1.3.4	Régression quantile	39

Chapitre 2 Détection du changement climatique et prévision de l'envasement par des copules **41**

	Introduction	42
2.1	Région d'étude	45
2.2	Données	47
2.2.1	Exploration graphique	48
2.2.2	Choix des périodes	49
2.2.3	Analyse des corrélations	50
2.2.4	Modélisation ARMA	50
2.3	Résultats	52
2.3.1	Analyse graphique des résidus	52
2.3.1.1	Nuages de points	52
2.3.1.2	K-plots	52
2.3.2	Estimation des paramètres	53
2.3.3	Comparaison des critères AIC	54
2.3.4	Résultats du test d'ajustement	54
2.3.5	Détection du changement climatique par les copules	54
2.3.6	Prévision de l'envasement	57
2.3.6.1	Restriction du choix des copules	57

2.3.6.2	Vérification de la qualité d'ajustement par simulation	57
2.3.6.3	Choix de la copule conditionnelle	58
2.3.6.4	Résultats de la prévision	59
Conclusion	61
Chapitre 3 Régression quantile par des copules-vignes pour la gestion d'un barrage en cascades		62
Introduction	63
3.1	Données	65
3.1.1	Région d'étude	65
3.1.2	Lois des distributions des variables	67
3.2	Modélisation par R-copule-vigne	70
3.2.1	La nouvelle R-vigne	70
3.2.2	La R-vigne classique	71
3.2.3	Test de Vuong	74
3.3	Régression quantile	76
3.3.1	Expression du quantile conditionnel	76
3.3.2	Résultats de la régression	77
3.4	Discussion	81
Conclusion	82
Conclusion générale et perspectives		83
Annexes		86
A	Suite des résultats du Chapitre 2	87
A.1	Tests d'indépendance	87
A.1.1	Test de Ljung-Box	87
A.1.2	Test de Box-Pierce	87
A.2	Tests de normalité	87
A.2.1	Test d'Anderson-Darling	87
A.2.2	Test de Kolmogorov-Smirnov	88
A.3	Estimation des données manquantes avec MICE	88
A.4	Statistiques descriptives, corrélations avec décalages et stationnarité des données	90
A.4.1	Statistiques descriptives	90
A.4.2	Corrélations avec décalages	90
A.4.3	Résultats de la stationnarité des séries	91

A.5	Résultats MLE et AIC des copules	92
A.5.1	Résultats MLE	92
A.5.2	Résultats AIC	93
A.6	Analyse temporelle des copules (suite)	94
B	Suite des résultats du Chapitre 3	97
B.1	Données	97
B.2	Représentation matricielle de la R-vigne	99
B.3	Scénarios de la régression quantile (suite)	101
C	Code R	103

Bibliographie

LISTE DES TABLEAUX

2.1	<i>Bathymétrie annuelle au barrage Mohammed V.</i>	47
2.2	<i>Mesures de corrélation entre les variables.</i>	50
2.3	<i>Modèles Autorégressif Moyenne Mobile (ARMA) retenus et résultats des tests de normalité et d'indépendance des résidus.</i>	51
2.4	<i>Résultat du test d'ajustement de Cramér-von Mises des copules retenues.</i>	55
2.5	<i>Valeurs prévues de l'envasement.</i>	60
3.1	<i>Codes, noms et surfaces de drainage des stations.</i>	66
3.2	<i>Description générale des fonctions de distribution utilisées dans cette étude.</i>	68
3.3	<i>Paramètres des distributions estimés par Estimation par la méthode du Maximum de Vraisemblance (MLE) et p-values du test Anderson-Darling.</i>	68
3.4	<i>Distributions marginales et résultats du test d'indépendance.</i>	69
3.5	<i>Copules bivariées et copules bivariées conditionnelles de la R-vigne à réseau réduit.</i>	72
3.6	<i>Comparaison des structures vignes de la variable volume.</i>	75
3.7	<i>Scénarios possibles en fonction de la gravité du quantile extrême de la station 15.</i>	78
A1	<i>Données de départ.</i>	89
A2	<i>Statistiques descriptives des variables.</i>	90
A3	<i>Résultats des tests de stationnarité des séries.</i>	92
A4	<i>Résultats de l'estimation des paramètres des copules par Estimation par la méthode du Maximum de Vraisemblance (MLE).</i>	93
A5	<i>Résultats des Critère d'Information d'Akaike (AIC) des copules pour les différentes périodes.</i>	94
B1	<i>Données des variables volume et débit de pointe des stations étudiées.</i>	97
C1	<i>Commandes et Packages R utilisés pour générer les résultats (1).</i>	103
C2	<i>Commandes et Packages R utilisés pour générer les résultats (2).</i>	104

TABLE DES FIGURES

1.1	<i>Densités des copules de Gauss et Student.</i>	9
1.2	<i>Relations entre les copules bivariées de valeurs extrêmes.</i>	14
2.1	<i>Quelques techniques de réalisation de relevés bathymétriques.</i>	44
2.2	<i>Barrage Mohammed V dans le bassin de la Moulouya.</i>	46
2.3	<i>Séries chronologiques des variables étudiées.</i>	48
2.4	<i>Nuages de points entre les couples de résidus de la période 2.</i>	53
2.5	<i>Exemples de K-plots obtenus.</i>	54
2.6	<i>Changements remarquables des contours des densités des copules.</i>	56
2.7	<i>Échantillons simulés à partir des copules ajustées.</i>	58
2.8	<i>Diagrammes en boîtes des modèles d'envasement.</i>	59
2.9	<i>Prévision de l'envasement en fonction du changement annuel de la température moyenne d'été.</i>	60
3.1	<i>Bassin et sous-bassins de la Saint-Jean.</i>	65
3.2	<i>Stations de contrôle sélectionnées pour illustrer la méthode R-vigne.</i>	67
3.3	<i>Premiers arbres de trois vignes basées sur la structure du bassin.</i>	70
3.4	<i>La R-vigne réseau réduit.</i>	71
3.5	<i>Contours des copules empiriques.</i>	72
3.6	<i>Contours des copules ajustées aux données.</i>	73
3.7	<i>Pair-plots, Taus de Kendall et histogrammes dans les cinq stations.</i>	73
3.8	<i>Tau D-vigne de la variable volume.</i>	74
3.9	<i>Tau R-vigne de la variable débit de pointe.</i>	74
3.10	<i>Principaux scénarios de la station 15 obtenus par régression quantile.</i>	79

3.11	<i>Densité des quantiles conditionnels à la station 15 en fonction des variations de probabilités à covariables fixes.</i>	80
A1	<i>Corrélations avec décalages.</i>	91
A2	<i>Contours des densités des copules obtenues entre l'apport en eau et les variables climatiques.</i>	95
A3	<i>Contours des densités des copules obtenues entre l'envasement et les variables climatiques.</i>	96
B1	<i>Séries chronologiques, histogrammes avec densités de lois théoriques et QQ-plots des variables étudiées.</i>	98
B2	<i>Étapes de construction de la matrice régulière de la R-vigne du bassin.</i>	100
B3	<i>Étapes de construction de la matrice régulière de la R-vigne du bassin (suite).</i>	101
B4	<i>Scénarios possibles pour la station 15 obtenus par régression quantile (suite).</i>	102

ACRONYMES

AIC Critère d'Information d' Akaike

ARMA Autorégressif Moyenne Mobile

BIC Critère d'Information Bayésien

CDF Fonction de Distribution Cumulative

GEV loi d'extremum généralisée

i.i.d indépendantes et identiquement distribuées

IM Indice Marginal

IO Indice Observé

MICE Multiple Imputation by Chained Equations

MLE Estimation par la méthode du Maximum de Vraisemblance

PIT transformées intégrales de probabilité

INTRODUCTION GÉNÉRALE

La modélisation et l'évaluation du risque sont des sujets qui requièrent de plus en plus d'attention dans le monde de la recherche scientifique. Elles consistent à représenter un système, incorporant généralement des distributions de probabilité, en utilisant des données historiques afin de comprendre la probabilité qu'un évènement à risque (naturel, économique ou financier) se produise et sa gravité potentielle.

La probabilité de dépassement donne la possibilité d'occurrence de ces risques à un niveau donné ou supérieur. Elle est utilisée pour planifier des dangers potentiels tels que les inondations, les sécheresses et la planification des niveaux de stockage de réservoirs en tenant compte des risques qui leur sont liés.

Classiquement, lorsque les variables aléatoires ne sont pas indépendantes, on utilise la loi gaussienne multivariée ou quelques lois bivariées comme la loi de Gumbel bivariée pour permettre une modélisation bi-/multi-dimensionnelle. Mais cette modélisation présente plusieurs limitations. D'abord, toutes les marginales doivent être de même loi, en l'occurrence des lois gaussiennes univariées pour la loi normale multivariée. En plus, cette dernière possède des queues légères et ne tient donc pas compte de la modélisation des valeurs extrêmes.

Un des objectifs de cette thèse est de trouver des distributions qui peuvent résoudre certains de ces problèmes pour l'estimation du risque d'apparition d'évènements extrêmes dans un cadre multivarié. L'utilisation des copules permet de résoudre partiellement cette problématique.

C'est dans ce cadre que notre thèse vient pour contribuer à la modélisation de certaines variables hydrologiques d'intérêt relatives aux barrages. Premièrement, nous proposons une méthode basée

sur les copules pour évaluer l'impact du changement climatique sur des variables hydrologiques liées à un barrage. Nous proposons également un modèle pour l'envasement qui servira à prédire ses quantités suivant différents scénarios de changement de température maximale. Deuxièmement, nous utilisons les copules-vignes pour concevoir un modèle multivarié des stations de contrôle d'un bassin hydrologique. Ce modèle servira à l'estimation des quantiles extrêmes des variables liées aux crues telles que le volume maximal et le débit de pointe.

Les copules ont été définies pour la première fois par *Sklar* [1959]. Elles sont connues pour leur capacité à modéliser le comportement joint de deux ou plusieurs variables en utilisant leurs distributions marginales. Il existe une certaine dépendance entre les variables aléatoires dont la copule capture les caractéristiques essentielles. Les avantages d'utilisation des copules pour modéliser les distributions jointes sont multiples. Elles fournissent une flexibilité dans le choix des marges arbitraires et de la structure de dépendance; la capacité d'extension à plus de deux variables; une analyse séparée des distributions marginales et de la structure de dépendance (voir *Singh* [2013]). Les copules ont été largement utilisées dans de nombreux domaines et ont fait l'objet de plusieurs études, notamment en finance, par exemple dans les travaux de *Palaro and Hotta* [2006], *Chu* [2011], *Dißmann et al.* [2013], *Ivanov et al.* [2017], *Fink et al.* [2017] et *Ait Hassou et al.* [2017] et dans notre domaine d'intérêt, l'hydrologie.

Les copules ont été utilisées dans *De Michele and Salvadori* [2003], par exemple, pour l'estimation des valeurs extrêmes en considérant les inondations comme un comportement joint de pic et de volume. Dans *Pujol et al.* [2008], une modélisation multivariée est réalisée par des fonctions copules pour étudier l'impact du changement climatique sur le régime des pluies extrêmes en examinant les changements régionaux dans des séries de précipitations extrêmes. Dans *Lee and Salas* [2008], on explore l'applicabilité du concept de la copule pour la simulation stochastique des ruissellements. Des fonctions paramétriques et non paramétriques sont utilisées pour ajuster la distribution des données observées d'origine et la structure de dépendance en série est modélisée avec des copules. Dans *Bárdossy and Li* [2008], la distribution conditionnelle des stations non observées est obtenue à partir de la copule conditionnelle et de la distribution marginale. Dans *Bárdossy and Pegram* [2009], les copules asymétriques sont utilisées pour modéliser la structure d'interdépendance spatiale entre le taux de précipitations et son occurrence en une seule relation. Dans *Samaniego et al.* [2010], les copules sont utilisées comme mesures de non similitude à travers la modélisation de la structure de corrélation de paires de séries chronologiques de ruissellement pour identifier la similarité des bassins et transférer les propriétés d'un bassin à un autre. Dans *Hao and Singh* [2012], la distribution jointe des débits mensuels adjacents est construite à l'aide de la méthode de la copule, tandis que la distribution marginale de l'écoulement fluvial pour chaque

mois est calculée à l'aide de la méthode de l'entropie. Dans *Hao and Singh* [2013], les structures de dépendance temporelle et spatiale sont imposées comme contraintes de la copule d'entropie maximale afin d'obtenir une simulation du ruissellement multi-site mensuel. Les flux mensuels sur différents sites sont ensuite générés par échantillonnage à partir de la distribution conditionnelle. Dans *Sugimoto et al.* [2016], l'asymétrie de la copule est utilisée pour capter les propriétés non symétriques des données de ruissellement et les distances de la copule sont utilisées pour identifier les changements dans un bassin en révélant la variabilité et l'interdépendance des structures de dépendance non révélées par la variance, la covariance et la corrélation.

Les copules sont de plus en plus utilisées dans les applications d'ingénierie des ressources hydriques. Certaines copules le sont encore plus, vu les différents avantages qu'elles présentent. Parmi ces avantages, nous en citons quelques uns discutés dans *Singh* [2013]. Les copules Archimédiennes sont préférables aux copules des autres familles grâce à leur forme simple, leur structure de dépendance et autres propriétés désirables. Les copules Meta-elliptiques, grâce à leur flexibilité, sont aussi utilisées pour la modélisation des données multivariées en hydrologie. Dans les années récentes, la copule de Plackett et les copules mixtes commencent aussi à être utilisées. Un dernier exemple est celui des copules empiriques. Elles sont utilisées quand les données sont analysées avec une distribution inconnue, les distributions empiriques des données peuvent donc être transformées en distributions marginales uniformes pour servir de distributions marginales à la copule empirique.

Pour tenir compte des dépendances multivariées, nous utilisons les copules-vignes. Une vigne est un outil graphique qui sert à modéliser les contraintes dans les distributions à grandes dimensions. Une vigne régulière est un cas particulier pour lequel les contraintes sont 2- dimensionnelles (2-D) ou 2-D conditionnelles. C'est une décomposition additive de l'information mutuelle dépendant uniquement de l'information mutuelle espérée de chaque côté (voir *Cooke et al.* [2010]). C'est dans *Joe* [1994] qu'il a été question de la première vigne régulière (R-vigne). Le but était d'étendre la copule bivariée de valeurs extrêmes à de plus grandes dimensions. En d'autres termes, il fallait trouver une classe de distributions n -variées à distributions marginales unidimensionnelles et $n(n - 1)$ paramètres de dépendance, avec $n - 1$ marges bivariées et n marges bivariées conditionnelles. La combinaison de la R-vigne avec les copules constitue un outil flexible pour la modélisation en haute dimension (voir *Aas et al.* [2009] et *Cooke et al.* [2010]). En effet, des constructions de pair-copules ont été comparées dans *Aas and Berg* [2009] avec d'autres modèles multivariés tels que les copules paramétriques n -dimensionnelles et les constructions Archimédiennes hiérarchiques. Il a été conclu par la supériorité des constructions pair-copules.

Les R-vignes ont de nombreuses propriétés remarquables détaillées dans *Bedford and Cooke*

[2002] et *Joe and Kurowicka* [2011]. Deux types particuliers de R-vignes sont la vigne canonique (C-) qui a une structure en étoile qui permet à une variable clé de gouverner les interactions dans la série de données et la "drawable" (D-) vigne qui a une structure de chemin. Les C- et D-vignes permettent d'introduire les variables dans l'ordre de priorité (voir *Xiong et al.* [2014]). Les R-vignes sont de plus en plus utilisées, principalement en finance et en climatologie. Par exemple, dans *Brechmann et al.* [2013], les vignes-copules sont utilisées pour analyser les interdépendances entre la banque et les assureurs sur le marché financier. Dans *Krouthén* [2015], on utilise les C-vignes pour analyser des actifs boursiers. Dans *Geidosch and Fischer* [2016], on démontre la supériorité des copules-vignes sur les copules conventionnelles lors de la modélisation de la structure de dépendance d'un portefeuille de crédits. Dans *Bevacqua et al.* [2017], on utilise les copules-vignes dans des stratégies de trading et d'allocation d'actifs. Dans *Fink et al.* [2017], on utilise des modèles R-vigne pour étudier l'existence de différents régimes globaux de dépendance. En climatologie, *Gyasi-Agyei* [2011] considère dans son travail une approche basée sur les copules pour modéliser la dépendance de facteurs multivariés de précipitations. Dans *Erhardt et al.* [2015], on prouve l'existence d'une relation entre les paramètres des copules-vignes et les distances des stations dans une étude de température moyenne quotidienne. Dans *Vernieuwe et al.* [2015], on utilise les copules-vignes pour incorporer toutes les dépendances pertinentes entre les variables de tempête. Dans *Gómez et al.* [2018], on utilise les copules-vignes pour modéliser la dépendance multivariée entre la décharge du glacier et d'autres variables météorologiques associées. Les R-vignes ont également été utilisées dans divers autres domaines, par exemple dans l'analyse de la sécurité aérienne (voir *Nápoles* [2009]) et pour capturer la dépendance spatiale et temporelle entre les centrales hydroélectriques (voir *Pereira et al.* [2016]). Nous pouvons trouver dans *Hernandez* [2015] un aperçu détaillé des applications des copules-vignes.

La suite de cette thèse se présente comme suit : Dans le Chapitre 1, nous exposons des outils statistiques qui constitueront notre environnement de travail. Nous y explicitons notamment les notions de copules et copules-vignes et leurs méthodes de construction respectives. Dans les Chapitres 2 et 3, nous détaillons les principaux résultats relatifs aux applications de notre méthode aux régions d'étude et données choisies. Nous concluons cette thèse en reprenant nos principales contributions ainsi que des perspectives d'étude. Du matériel supplémentaire est disponible en fin de ce document dans la partie "Annexes".

CHAPITRE 1

THÉORIE DES COPULES ET COPULES-VIGNES

« When there is a need to explain a general dependence structure then any field looking for dependence measurements can consider copula modelling. »

Alina [2018]

« All families of multivariate models and their associated statistical techniques (inference, testing, simulation, etc) potentially have to be revisited under a copula point of view for theoretical and practical reasons. »

Fermanian [2017]

1.1 La théorie des copules

Les copules empiriques feront l'objet de la sous-section 1.1.2. Les copules elliptiques et Archimédiennes seront traitées dans la sous-section 1.1.3. Mais d'abord, commençons par introduire le théorème le plus important relatif aux copules, **le théorème de Sklar**.

1.1.1 Théorème de Sklar

Le théorème de Sklar est le théorème pilier de la théorie des copules. Il vient pour élucider le rôle que joue la copule dans la relation entre les fonctions de distributions multivariées et leurs distributions marginales univariées. Dans ce paragraphe, nous faisons un rappel de quelques définitions et théorèmes de base qui constitueront notre environnement de travail. Ces outils méthodologiques se trouvent dans *Nelsen* [2007], *Joe* [2014] et *Czado* [2019]. Les démonstrations des théorèmes se trouvent essentiellement dans *Nelsen* [2007] et sont reprises en cas de nécessité. Pour les besoins de notre étude, toutes les définitions et théorèmes ne seront donnés que dans le cas bivarié.

Définition 1.1. Une fonction de distribution est une fonction F de domaine $\bar{\mathbb{R}} = \mathbb{R} \cup \{-\infty, \infty\}$ telle que

1. F est non décroissante.
2. $F(-\infty) = 0$ et $F(\infty) = 1$.

Définition 1.2. Une fonction de distribution jointe est une fonction H de domaine $\bar{\mathbb{R}}^2$ telle que

1. H est 2-croissante.
2. $H(x, -\infty) = H(-\infty, y) = 0$ et $H(\infty, \infty) = 1$.

Nous pouvons maintenant introduire la notion de copule.

Définition 1.3. Une copule bidimensionnelle est une fonction C de $[0, 1]^2$ dans $[0, 1]$ ayant les propriétés suivantes :

1. pour tout u et v dans $[0, 1]$ nous avons :

$$C(u, 0) = 0 = C(0, v); C(u, 1) = u \text{ et } C(1, v) = v.$$

2. pour tout u_1, u_2, v_1 et v_2 dans $[0, 1]$ tel que $u_1 \leq u_2$ et $v_1 \leq v_2$, C est 2-croissante, i.e :

$$C(u_2, v_2) - C(u_2, v_1) - C(u_1, v_2) + C(u_1, v_1) \geq 0.$$

Les propriétés (1) et (2) définissent une copule bivariee comme une fonction de distribution sur $[0, 1]^2$ avec des distributions marginales uniformement distribuées (voir *Nelsen* [2007] et *Eschenburg* [2013]), c-à-d. pour des variables aléatoires U_1, U_2 ayant pour distribution jointe une certaine copule C , nous avons $U_1 \sim U[0, 1]$ et $U_2 \sim U[0, 1]$. L'existence de la copule est garantie par le théorème de Sklar.

Théorème 1.1 (*Sklar* [1959]). *Soit H une fonction de distribution jointe avec des distributions marginales F et G . Il existe alors une copule C telle que :*

$$H(x, y) = C(F(x), G(y)), \forall x, y \in \mathbb{R}. \quad (1.1)$$

Inversement, si C est une copule et F et G sont des fonctions de distribution marginales univariées alors H est une fonction de distribution jointe avec des marges F et G : Si F et G sont continues, alors on peut définir C de manière unique par

$$C(u, v) = H(F^{-1}(u), G^{-1}(v)). \quad (1.2)$$

Ceci équivaut à dire que pour identifier la distribution jointe $H(x, y)$ d'un échantillon aléatoire $(X_1, Y_1) \dots (X_n, Y_n)$ issu de la paire de variables aléatoires continues (X, Y) , le théorème de Sklar assure l'existence d'une copule unique telle que $H(x, y) = C(F(x), G(y))$.

1.1.2 La copule empirique

La copule empirique est la forme de dépendance la plus intuitive à laquelle nous pouvons penser. Elle a été introduite dans le but de faire des inférences sur la copule sous-jacente à un échantillon aléatoire $(X_1, Y_1) \dots (X_n, Y_n)$ issu d'une paire de variables aléatoires continues (X, Y) . En utilisant une approximation basée sur les données de copules $u = F(x)$ et $v = G(y)$ correspondant à ces observations, la copule empirique est construite comme suit (voir *Czado* [2019]) :

$$\hat{C}(u_1, u_2) := \frac{1}{n+1} \sum_{i=1}^n 1_{\{u_{1i} \leq u_1, u_{2i} \leq u_2\}}, \quad (1.3)$$

pour tout $0 \leq u_1, u_2 \leq 1$.

Remarque 1.1. *Ici, \hat{C} fait référence à une estimation de la copule empirique. C'est le seul cas où cette notation sera utilisée dans ce sens. Pour la suite, \hat{C} désignera la copule de survie.*

1.1.3 Familles des copules

Il existe plusieurs familles de copules qui peuvent être classées en trois grandes familles (voir *Nelsen* [2007] et *Czado* [2019]). La première est la famille des copules elliptiques dont nous donnerons l'exemple de la copule Gaussienne et celle de Student. La seconde est celle des copules Archimédiennes pouvant être obtenues à partir de fonctions réelles dites *fonctions génératrices*. Nous utiliserons les copules de Clayton, Gumbel, Frank et Joe comme exemples de cette famille. La troisième approche découle des extensions de la théorie des valeurs extrêmes univariées à des dimensions supérieures dont nous donnerons l'exemple de la copule de Tawn (voir *Tawn* [1988]). Pour une revue complète des familles des copules voir *Hutchinson* [1990], *Joe* [1996] et *Nelsen* [2003, 2007].

1.1.3.1 Famille elliptique

Les copules elliptiques sont des copules obtenues à partir de distributions elliptiques. Nous reprenons ici l'exemple de la copule gaussienne et de la copule de Student comme détaillé dans *Czado* [2019].

La copule gaussienne n'a pas de dépendance de queue ou d'asymétrie de queue. Grâce à son degré de liberté, la copule de Student permet de tenir compte des distributions à queues épaisses et la modélisation par une telle copule convient lorsque la propriété de symétrie est maintenue. Cependant, ces deux copules n'ont pas de forme de Fonction de Distribution Cumulative (CDF) connue ce qui peut être un problème pour les applications avec des données discrètes multivariées (voir *Joe* [2014] et *Czado* [2019]). La copule gaussienne bivariée est construite en utilisant une distribution normale bivariée de moyenne nulle, de variance 1 et de corrélation ρ notée $\Phi_2(\cdot, \cdot; \rho)$ en appliquant la version inverse du théorème 1.1, page 7 suivante :

$$C(u_1, u_2) = F(F_1^{-1}(u_1), F_2^{-1}(u_2)), \quad (1.4)$$

ce qui nous donne

$$C(u_1, u_2; \rho) = \Phi_2(\Phi^{-1}(u_1), \Phi^{-1}(u_2); \rho), \quad (1.5)$$

$\Phi(\cdot)$ étant la fonction de distribution normale standard $N(0, 1)$.

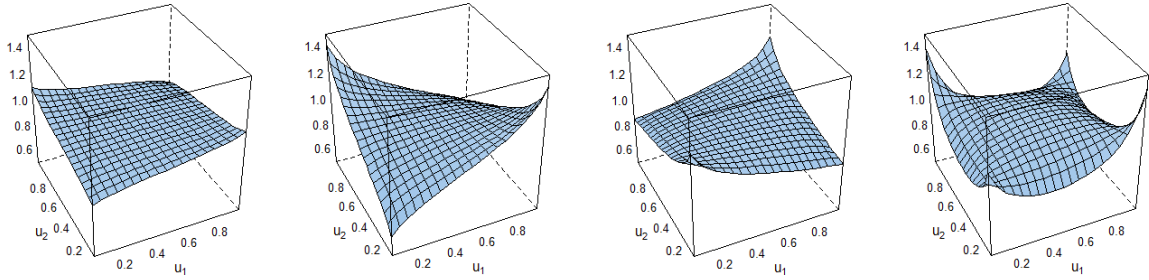
La copule de Student bivariée est construite en utilisant la distribution Student t bivariée à ν degrés de liberté, à moyenne nulle et à corrélation ρ en intégrant la densité de copule donnée dans l'équation 1.4, page 8 suivante :

$$c(u_1, u_2; \rho) = \frac{f(F_1^{-1}(u_1), F_2^{-1}(u_2))}{f_1(F_1^{-1}(u_1))f_2(F_2^{-1}(u_1))}, \quad (1.6)$$

ce qui nous donne pour la densité bivariee de Student notée $t(x_1, x_2; \mathbf{v}, \rho)$

$$C(u_1, u_2; \mathbf{v}, \rho) = \int_0^{u_1} \int_0^{u_2} \frac{t(T_{\mathbf{v}}^{-1}(v_1), T_{\mathbf{v}}^{-1}(v_2); \mathbf{v}, \rho)}{t_{\mathbf{v}}(T_{\mathbf{v}}^{-1}(v_1)), t_{\mathbf{v}}(T_{\mathbf{v}}^{-1}(v_2))} dv_1 dv_2. \quad (1.7)$$

Nous représentons dans la Figure 1.1 les densités de ces deux copules quand $\rho = 0.2$ et $\rho = 0.8$.



(a) Gauss : $\rho = 0.2$ (b) Gauss : $\rho = 0.8$ (c) Student : $\rho = 0.2, \nu = 8$ (d) Student : $\rho = 0.8, \nu = 8$

FIGURE 1.1 – Densités des copules de Gauss et Student.

Remarque 1.2. *Les copules relevant de la famille elliptique ne seront pas utilisées dans notre thèse. Nous utilisons des modèles basés sur des copules Archimédiennes ou de valeurs extrêmes. Cette préférence sera justifiée par certaines propriétés qui feront l'objet des paragraphes suivants.*

1.1.3.2 Famille Archimédienne

Pour la famille Archimédienne, les copules sont obtenues à partir d'une fonction génératrice.

Définition 1.4. *Czado [2019] (Copules Archimédiennes bivariées) Soit Ω l'ensemble de toutes les fonctions continues, strictement décroissantes et convexes $\phi : I \rightarrow [0, \infty]$ avec $\phi(1) = 0$. Soit $\phi \in \Omega$, alors*

$$C(u, v) = \phi^{[-1]}(\phi(u) + \phi(v)) \quad (1.8)$$

est une copule. C est dite une copule Archimédienne bivariée de générateur ϕ . Quand $\phi(0) = \infty$, le générateur est dit strict. Ici $\phi^{[-1]}$ est le pseudo-inverse de ϕ défini par $\phi^{[-1]} : [0, \infty] \rightarrow [0, 1]$ avec

$$\phi^{[-1]}(t) = \begin{cases} \phi^{-1}(t), & 0 \leq t \leq \phi(0) \\ 0, & \phi(0) \leq t \leq \infty \end{cases}$$

Vérifions que C est une copule.

Démonstration. Pour la condition (1) de la définition 1.3, nous avons

$$C(u, 0) = \phi^{[-1]}(\phi(u) + \phi(0)) = 0$$

et

$$C(u, 1) = \phi^{[-1]}(\phi(u) + \phi(1)) = \phi^{[-1]}(\phi(u)) = u.$$

Et par symétrie, nous avons

$$C(0, v) = 0 \text{ et } C(1, v) = v.$$

Pour la condition (2), soit t dans \mathbf{I} tel que

$$C(t, v_2) = v_1, \text{ où } \phi(v_2) + \phi(t) = \phi(v_1).$$

Alors

$$\begin{aligned} C(u_2, v_1) - C(u_1, v_1) &= \phi^{[-1]}(\phi(u_2) + \phi(v_1)) - \phi^{[-1]}(\phi(u_1) + \phi(v_1)) \\ &= \phi^{[-1]}(\phi(u_2) + \phi(v_2) + \phi(t)) - \phi^{[-1]}(\phi(u_1) + \phi(v_2) + \phi(t)) \\ &= C(C(u_2, v_2), t) - C(C(u_1, v_2), t) \\ &\leq C(u_2, v_2) - C(u_1, v_2), \end{aligned}$$

d'où C est 2-croissante. □

Nous citons ci-après quelques propriétés algébriques¹ des copules Archimédiennes.

Théorème 1.2 (Nelsen [2007]). *Soit C une copule Archimédienne de générateur ϕ . Alors*

1. C est symétrique ;
2. C est associative ;
3. si $c > 0$ est une constante, alors $c\phi$ est aussi un générateur de C .

Démonstration. Pour la propriété (1), pour (u, v) dans $[0, 1]^2$ nous avons

$$C(u, v) = \phi^{[-1]}(\phi(u) + \phi(v)) = \phi^{[-1]}(\phi(v) + \phi(u)) = C(v, u).$$

1. Les copules Archimédiennes possèdent des propriétés d'échangeabilité, la symétrie en l'occurrence, en contrepartie des copules périodiques ou mixtes, qui sont, elles, asymétriques Singh [2013].

Pour la propriété (2), pour (u, v, w) dans $[0, 1]^3$, nous avons

$$\begin{aligned}
C(C(u, v), w) &= \phi^{[-1]} \left(\phi \left(\phi^{[-1]}(\phi(u) + \phi(v)) \right) + \phi(w) \right) \\
&= \phi^{[-1]} (\phi(u) + \phi(v) + \phi(w)) \\
&= \phi^{[-1]} \left(\phi(u) + \phi \left(\phi^{[-1]}(\phi(v) + \phi(w)) \right) \right) \\
&= \phi^{[-1]} (\phi(u) + \phi(C(v, w))) \\
&= C(u, C(v, w))
\end{aligned}$$

Pour la propriété (3), nous avons pour une constante $c > 0$

$$\begin{aligned}
(c\phi)^{[-1]}((c\phi)(u) + (c\phi)(v)) &= \frac{1}{c} \phi^{[-1]}[c(\phi(u) + \phi(v))] \\
&= \phi^{[-1]}(\phi(u) + \phi(v)) \\
&= C(u, v).
\end{aligned}$$

□

Nous donnons ci-après des exemples d'obtention de distributions de copules Archimédiennes à partir de leurs fonctions génératrices données dans *Nelsen* [2007].

Exemple 1.1. Pour la copule de Clayton, la fonction génératrice est $\phi_C(t) = \frac{(t^{-\theta}-1)}{\theta}$ où $\theta > 0$, l'inverse de cette fonction est la fonction $\phi^{-1}(t) = (1 + \theta t)^{\frac{-1}{\theta}}$, nous avons

$$\begin{aligned}
C_C(u, v) &= \phi_C^{-1} (\phi_C(u) + \phi_C(v)) \\
&= (1 + \theta (\phi_C(u) + \phi_C(v)))^{\frac{-1}{\theta}} \\
&= \left(1 + \theta \left(\frac{(u^{-\theta} - 1)}{\theta} + \frac{(v^{-\theta} - 1)}{\theta} \right) \right)^{\frac{-1}{\theta}} \\
&= \left(u^{-\theta} + v^{-\theta} - 1 \right)^{\frac{-1}{\theta}} \tag{1.9}
\end{aligned}$$

Exemple 1.2. Pour la copule de Frank, la fonction génératrice est $\phi_F(t) = -\ln \left[\frac{e^{-\theta t} - 1}{e^{-\theta} - 1} \right]$, dont

l'inverse est $\phi_F^{-1}(t) = \frac{-1}{\theta} \ln [1 - e^{-t}(1 - e^{-\theta})]$, nous avons

$$\begin{aligned}
C_F(u, v) &= \phi_F^{-1}(\phi_F(u) + \phi_F(v)) \\
&= \frac{-1}{\theta} \ln \left[1 - \exp[-(\phi_F(u) + \phi_F(v))](1 - e^{-\theta}) \right] \\
&= \frac{-1}{\theta} \ln \left[1 - \exp \left[\ln \left(\frac{e^{-\theta u} - 1}{e^{-\theta} - 1} \right) + \ln \left(\frac{e^{-\theta v} - 1}{e^{-\theta} - 1} \right) \right] (1 - e^{-\theta}) \right] \\
&= \frac{-1}{\theta} \ln \left[1 - \exp \left[\ln \left(\frac{(e^{-\theta u} - 1)(e^{-\theta v} - 1)}{(e^{-\theta} - 1)^2} \right) \right] (1 - e^{-\theta}) \right] \\
&= \frac{-1}{\theta} \ln \left[1 - \frac{(e^{-\theta u} - 1)(e^{-\theta v} - 1)}{(e^{-\theta} - 1)^2} (1 - e^{-\theta}) \right] \\
&= \frac{-1}{\theta} \ln \left[1 + \frac{(e^{-\theta u} - 1)(e^{-\theta v} - 1)}{e^{-\theta} - 1} \right]
\end{aligned} \tag{1.10}$$

Exemple 1.3. La copule de Gumbel permet de mesurer la dépendance des événements les plus rares et possède un paramètre de dépendance unique. Elle capte la dépendance dans l'extrémité supérieure de la distribution *Pujol et al.* [2008]. C'est la copule à valeurs extrêmes la plus utilisée. Sa fonction génératrice est donnée par $\phi_G(t) = (-\ln t)^\theta$ où $1 \leq \theta < \infty$, dont l'inverse est $\phi_G^{-1}(t) = \exp(-t^{\frac{1}{\theta}})$, nous avons

$$\begin{aligned}
C_G(u, v) &= \phi_G^{-1}(\phi_G(u) + \phi_G(v)) \\
&= \exp\left(-[\phi_G(u) + \phi_G(v)]^{\frac{1}{\theta}}\right) \\
&= \exp\left(-\left[(-\ln u)^\theta + (-\ln v)^\theta\right]^{\frac{1}{\theta}}\right)
\end{aligned} \tag{1.11}$$

Exemple 1.4. La copule de Joe a un paramètre de dépendance unique qui peut présenter un paramètre de dépendance spécifique. Sa fonction génératrice est donnée par $\phi_J(t) = -\ln [1 - (1 - t)^\theta]$ où $\theta \geq 1$, dont l'inverse est $\phi_J^{-1}(t) = 1 - \left((1 - e^{-t})^{\frac{1}{\theta}}\right)$, nous avons

$$\begin{aligned}
C_J(u, v) &= \phi_J^{-1}(\phi_J(u) + \phi_J(v)) \\
&= 1 - [1 - \exp(\phi_J(u) + \phi_J(v))]^{\frac{1}{\theta}} \\
&= 1 - \left[1 - \exp\left(\ln [1 - (1 - u)^\theta] + \ln [1 - (1 - v)^\theta]\right)\right]^{\frac{1}{\theta}} \\
&= 1 - \left[(1 - u)^\theta + (1 - v)^\theta + (1 - u)^\theta (1 - v)^\theta\right]^{\frac{1}{\theta}}
\end{aligned} \tag{1.12}$$

D'autres copules Archimédiennes à deux paramètres telles que Clayton-Gumbel (dite BB1), Joe-Gumbel (dite BB6), Joe-Clayton (dite BB7) et Joe-Frank (dite BB8) sont à structure plus

flexible et permettent des coefficients de dépendance non nuls dans les queues supérieures et inférieures (voir *Joe* [2014]). Mais plusieurs copules n'ont pas de dépendance dans les queues ce qui veut dire qu'elle sous-estiment le risque d'évènements extrêmes.

1.1.3.3 Copules de valeurs extrêmes

Les copules de valeurs extrêmes sont des copules spéciales que nous utilisons dans notre étude. Elles sont très appropriées pour capturer le co-mouvement des queues des fonctions de distribution multivariées. Ces copules mesurent la probabilité d'une occurrence simultanée de valeurs extrêmes de deux ou plusieurs variables aléatoires et conviennent donc à la modélisation de phénomènes de valeurs extrêmes dans les domaines financier, économique et environnemental. Nous donnerons ci-après la définition d'une copule de valeurs extrêmes telle que reprise par *Eschenburg* [2013] :

Définition 1.5 (*Pickands* [1981]). Soit $A : [0, 1] \rightarrow [\frac{1}{2}, 1]$ une fonction convexe satisfaisant

$$\max(w, 1 - w) \leq A(w) \leq 1, \quad (1.13)$$

pour tout $w \in [0, 1]$. La copule de valeurs extrêmes est définie par :

$$C(u, v) = \exp \left[\log(uv) A \left(\frac{\log v}{\log(uv)} \right) \right] \quad (1.14)$$

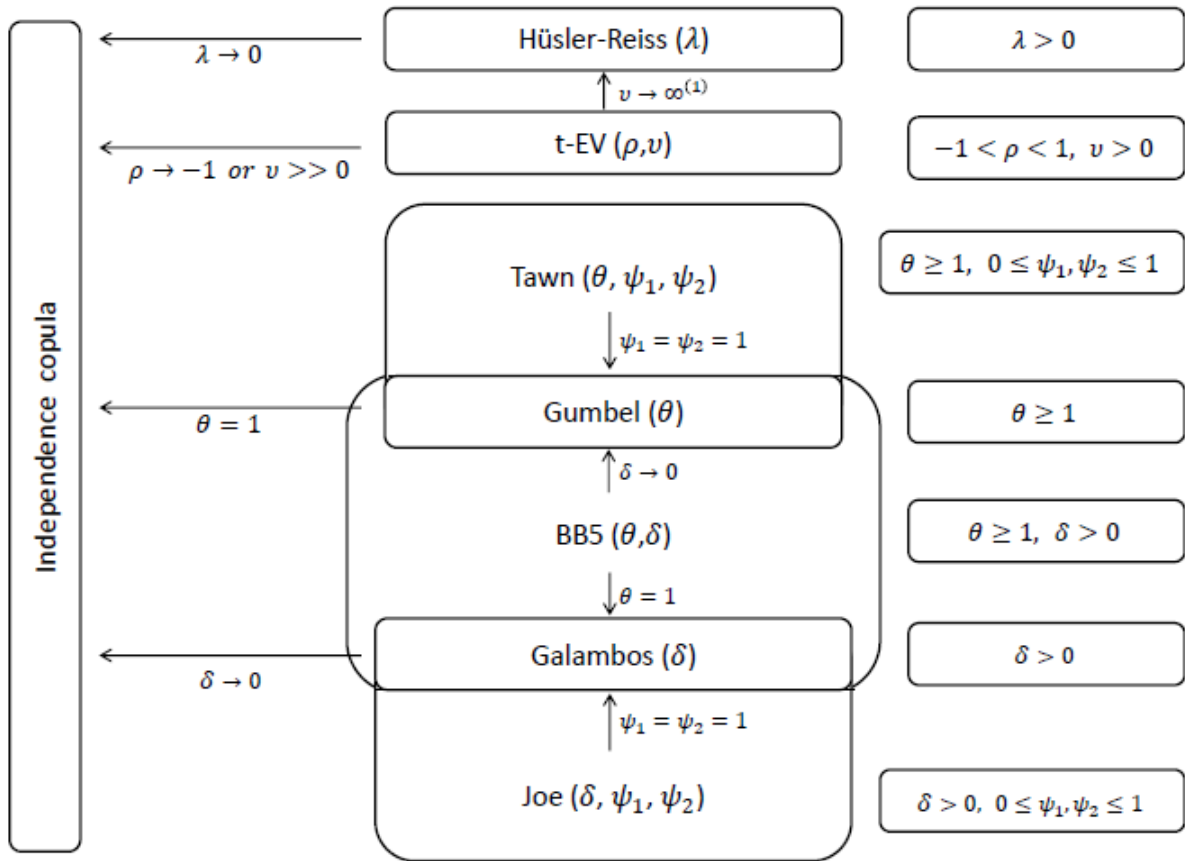
Nous trouvons dans *Czado* [2019] une autre définition des copules de valeurs extrêmes

Définition 1.6. Une copule bivariée C est dite *copule de valeurs extrêmes* s'il existe une copule bivariée C_X telle que pour $n \rightarrow \infty$ nous avons,

$$[C_X(u_1^{1/n}, u_2^{1/n})]^n \rightarrow C(u_1, u_2) \forall (u_1, u_2) \in [0, 1]^2. \quad (1.15)$$

La copule C_X est dite être dans le domaine d'attraction de la copule C .

Les paramètres de dépendance des copules de valeurs extrêmes déterminent les liens entre ces copules et permettent le passage d'une copule à une autre, comme on peut le voir sur la Figure 1.2. Pour les copules de Joe et Tawn (définie dans le paragraphe suivant), les mesures de dépendance atteignent leur maximum quand les deux paramètres d'asymétrie atteignent leur maximum ($\psi_1 = \psi_2 = 1$), une grande asymétrie signifie donc une très grande dépendance. Ces deux copules ont des paramètres de dépendance définis sur $[0, 1]$, les copules de Galambos et Gumbel (respectivement) en sont des cas particuliers lorsque les deux paramètres atteignent leurs valeurs maximales (voir *Coles and Tawn* [1991]). Elles sont particulièrement présentes dans la littérature des copules de valeurs extrêmes pour leur connexion avec la copule de Gumbel. Leurs fonctions de



Source : Eschenburg [2013]

FIGURE 1.2 – Relations entre les copules bivariées de valeurs extrêmes.

dépendance de Pickands ne peuvent être distinguées l'une de l'autre, elles sont presque identiques *Coles and Tawn* [1991]. En plus, le paramètre θ dans la copule de Tawn a la même interprétation que dans la copule de Gumbel et reste le facteur principal pour déterminer le degré de dépendance (voir *Tawn* [1988]).

Une copule très utilisée appartenant à la famille des copules de valeur extrême est la copule de Tawn. Aussi connue sous le nom de modèle logistique asymétrique, elle a été obtenue par *Tawn* [1988] en ajoutant 2 paramètres à la copule de Gumbel pour remédier à l'hypothèse de symétrie pas toujours établie $C(u, v) = C(v, u)$. La copule de Tawn étant l'équivalent non symétrique de la copule de Gumbel est donc souvent recommandée pour la modélisation des comportements joints asymétriques (voir *Pujol et al.* [2008]). La copule de Tawn permet de modéliser une probabilité d'évènements extrêmes joints supérieure à celle modélisée par la copule gaussienne (voir *Yoo and Cho* [2018]). Pour permettre une modélisation plus simple de cette extension à trois paramètres, *Schepsmeier et al.* [2015] a élaboré l'expression de deux versions de la copule de Tawn avec deux paramètres chacune. Chaque type a l'un des paramètres d'asymétrie fixé à 1, de sorte que la densité

de copule correspondante est soit asymétrique à gauche ou à droite (par rapport à la diagonale principale). Ces deux nouvelles copules ont été nommées Tawn type 1 (Tawn T1) et Tawn type 2 (Tawn T2) et sont définies comme suit :

Définition 1.7 (*Schepsmeier et al. [2015]*). Les fonctions génératrices des copules Tawn T1 et T2 sont respectivement :

$$A_1(w) = 1 - (\theta + \phi)w + \theta w^2 + \phi w^3, \quad (1.16)$$

$$A_2(w) = (1 - \phi_1)(1 - w) + (1 - \phi_2)w + [(\phi_1 w)^{\frac{1}{\theta}} + (\phi_2(1 - w))^{\frac{1}{\theta}}]^{\theta}, \quad (1.17)$$

où w, ϕ, ϕ_1 , et ϕ_2 sont dans $[0, 1]$ et $\theta \in [0, \infty]$.

Ci-après, nous détaillons la façon d'obtenir les fonctions de distribution des copules Tawn T1 et Tawn T2 à partir de leurs fonctions génératrices.

Exemple 1.5. Des équations 1.14 et 1.16, nous obtenons la distribution de la copule Tawn T1 :

$$\begin{aligned} C_{T1}(u, v) &= \exp \left[\log(uv) - (\theta + \phi) \log(v) + \theta \frac{(\log(v))^2}{\log(uv)} + \phi \frac{(\log(v))^3}{(\log(uv))^2} \right] \\ &= (uv - v^{(\theta+\phi)}) \times \exp \left[\frac{(\log(v))^2}{\log(uv)} \left(\theta + \phi \frac{\log(v)}{\log(uv)} \right) \right] \end{aligned} \quad (1.18)$$

Exemple 1.6. Des équations 1.14 et 1.17, nous obtenons la distribution de la copule Tawn T2 :

$$\begin{aligned} C_{T2}(u, v) &= \exp \left[\log(uv) \left((1 - \phi_1) \left(1 - \frac{\log(v)}{\log(uv)} \right) + (1 - \phi_2) \frac{\log(v)}{\log(uv)} + \left[\left(\phi_1 \frac{\log(v)}{\log(uv)} \right)^{\frac{1}{\theta}} + \left(\phi_2 \left(1 - \frac{\log(v)}{\log(uv)} \right) \right)^{\frac{1}{\theta}} \right]^{\theta} \right) \right] \\ &= \exp \left[\log(uv) \left((1 - \phi_1) \left(\frac{\log(u)}{\log(uv)} \right) + (1 - \phi_2) \frac{\log(v)}{\log(uv)} + \left[\left(\phi_1 \frac{\log(v)}{\log(uv)} \right)^{\frac{1}{\theta}} + \left(\phi_2 \left(\frac{\log(u)}{\log(uv)} \right) \right)^{\frac{1}{\theta}} \right]^{\theta} \right) \right] \\ &= \exp \left[(1 - \phi_1) \log(u) + (1 - \phi_2) \log(v) + \left[(\log(uv))^{\frac{1}{\theta}} \left(\left(\phi_1 \frac{\log(v)}{\log(uv)} \right)^{\frac{1}{\theta}} + \left(\phi_2 \left(\frac{\log(u)}{\log(uv)} \right) \right)^{\frac{1}{\theta}} \right) \right]^{\theta} \right] \\ &= u^{(1-\phi_1)} v^{(1-\phi_2)} \exp \left[(\phi_1 \log(v))^{\frac{1}{\theta}} + (\phi_2 \log(u))^{\frac{1}{\theta}} \right]^{\theta} \end{aligned} \quad (1.19)$$

1.1.4 Copules de survie et rotations

Pour étendre la plage de dépendance, nous utilisons des rotations anti-horaire de la densité de copule $c(., .)$ (voir *Czado [2019]*). La rotation des copules par 90 et 270 degrés permet la modélisation de la dépendance négative qui n'est pas possible avec les versions sans rotation. Les paramètres de dépendance des copules à 90 et 270 degrés sont négatifs (voir *Joe [2014]*). Ci-après, on rappelle comment obtenir la rotation des copules par 180 degrés (appelée copule de survie) et les autres rotations des copules.

Définition 1.8 (*Nelsen* [2007]). Une copule de survie est une copule pivotée de 180 degrés, c'est une fonction $\hat{C} : I^2 \rightarrow I$ définie par :

$$\hat{C}(u, v) = u + v - 1 + C(1 - u, 1 - v) = \mathbb{P}[U > u, V > v], \quad (1.20)$$

pour tout $(u, v) \in I^2$.

Démonstration. Soit H la fonction de distribution jointe introduite précédemment dans le théorème de Sklar 1.1. En utilisant les notations de *Nelsen* [2007] pour les fonctions de survie marginales $\bar{F}(x) = P[X > x] = 1 - F(x)$, $\bar{G}(y) = P[Y > y] = 1 - G(y)$ et pour la fonction de survie jointe $\bar{H}(x, y) = P[X > x, Y > y]$, nous avons

$$\begin{aligned} \bar{H}(x, y) &= P[X > x] + P[Y > y] - P[X > x \cup Y > y] \\ &= 1 - F(x) + 1 - G(y) - P[\overline{X \leq x \cup Y \leq y}] \\ &= 1 - F(x) - G(y) + 1 - P[\overline{X \leq x, Y \leq y}] \\ &= 1 - F(x) - G(y) + P[X \leq x, Y \leq y] \\ &= \bar{F}(x) + \bar{G}(y) - 1 + H(x, y) \\ &= \bar{F}(x) + \bar{G}(y) - 1 + C(F(x), G(y)) \\ &= \bar{F}(x) + \bar{G}(y) - 1 + C(1 - \bar{F}(x), 1 - \bar{G}(y)) \end{aligned}$$

Il suffit alors de poser $u = \bar{F}(x)$ et $v = \bar{G}(y)$ pour obtenir la définition. \square

Les deux autres versions pivotées des copules sont à 90 et 270 degrés. Leurs fonctions de distribution sont respectivement :

$$C_{90}(u, v) = v - C(1 - u, v), \quad (1.21)$$

$$C_{270}(u, v) = u - C(u, 1 - v) \quad (1.22)$$

Nous donnons ci-après les expressions de ces trois rotations pour les copules utilisées dans cette partie de la thèse.

Exemple 1.7. Des équations 1.9 et 1.21, puis 1.9 et 1.20, puis 1.9 et 1.22 nous obtenons respecti-

vement la rotation à 90° , la copule de survie et la rotation à 270° de la copule de Clayton :

$$C_{r90C}(u, v) = v - ((1-u)^{-\delta} + v^{-\delta} - 1)^{-\frac{1}{\delta}} \quad (1.23)$$

$$\hat{C}_C(u, v) = u + v - 1 + ((1-u)^{-\delta} + (1-v)^{-\delta} - 1)^{-\frac{1}{\delta}} \quad (1.24)$$

$$C_{r270C}(u, v) = u - ((u)^{-\delta} + (1-v)^{-\delta} - 1)^{-\frac{1}{\delta}}. \quad (1.25)$$

Exemple 1.8. Des équations 1.12 et 1.21, puis 1.12 et 1.20, puis 1.12 et 1.22 nous obtenons respectivement la rotation à 90° , la copule de survie et la rotation à 270° de la copule de Joe :

$$C_{r90J}(u, v) = v - 1 + (u^\theta + (1-v)^\theta - u^\theta(1-v)^\theta)^{\frac{1}{\theta}} \quad (1.26)$$

$$\hat{C}_J(u, v) = u + v - (u^\theta + v^\theta - u^\theta v^\theta)^{\frac{1}{\theta}} \quad (1.27)$$

$$C_{r270J}(u, v) = u - 1 + ((1-u)^\theta + v^\theta - (1-u)^\theta v^\theta)^{\frac{1}{\theta}}. \quad (1.28)$$

Exemple 1.9. Des équations 1.11 et 1.21, puis 1.11 et 1.20, puis 1.11 et 1.22 nous obtenons respectivement la rotation à 90° , la copule de survie et la rotation à 270° de la copule de Gumbel :

$$C_{r90G}(u, v) = v - \exp - [(-\ln(1-u))^\delta + (-\ln(v))^\delta]^{\frac{1}{\delta}} \quad (1.29)$$

$$\hat{C}_G(u, v) = u + v - 1 + \exp - [(-\ln(1-u))^\delta + (-\ln(1-v))^\delta]^{\frac{1}{\delta}} \quad (1.30)$$

$$C_{r270G}(u, v) = u - \exp - [(-\ln(u))^\delta + (-\ln(1-v))^\delta]^{\frac{1}{\delta}}. \quad (1.31)$$

Exemple 1.10. Des équations 1.18 et 1.21, puis 1.18 et 1.20, puis 1.18 et 1.22 nous obtenons respectivement la rotation à 90° , la copule de survie et la rotation à 270° de la copule Tawn Type 1 :

$$C_{r90T1}(u, v) = v - \left((1-u)v - v^{(\theta+\phi)} \right) \times \exp \left[\frac{(\log(v))^2}{\log((1-u)v)} \left(\theta + \phi \frac{\log(v)}{\log((1-u)v)} \right) \right] \quad (1.32)$$

$$\hat{C}_{T1}(u, v) = u + v - 1 + \left((1-u)(1-v) - (1-v)^{(\theta+\phi)} \right) \times \exp \left[\frac{(\log(1-v))^2}{\log((1-u)(1-v))} \left(\theta + \phi \frac{\log(1-v)}{\log((1-u)(1-v))} \right) \right] \quad (1.33)$$

$$C_{r270T1}(u, v) = u - \left(u(1-v) - (1-v)^{(\theta+\phi)} \right) \times \exp \left[\frac{(\log(1-v))^2}{\log(u(1-v))} \left(\theta + \phi \frac{\log(1-v)}{\log(u(1-v))} \right) \right]. \quad (1.34)$$

Exemple 1.11. De la même façon, nous obtenons des équations 1.19 et 1.21, puis 1.19 et 1.20, puis

1.19 et 1.22 la rotation à 90° , la copule de survie et la rotation à 270° de la copule Tawn Type 2 :

$$C_{r90T2}(u, v) = v - (1 - u)^{(1-\phi_1)} v^{(1-\phi_2)} \exp \left[(\phi_1 \log(v))^{\frac{1}{\theta}} + (\phi_2 \log(1 - u))^{\frac{1}{\theta}} \right]^\theta \quad (1.35)$$

$$\hat{C}_{T2}(u, v) = u + v - 1 + (1 - u)^{(1-\phi_1)} (1 - v)^{(1-\phi_2)} \exp \left[(\phi_1 \log(1 - v))^{\frac{1}{\theta}} + (\phi_2 \log(1 - u))^{\frac{1}{\theta}} \right]^\theta \quad (1.36)$$

$$C_{r270T2}(u, v) = u - u^{(1-\phi_1)} (1 - v)^{(1-\phi_2)} \exp \left[(\phi_1 \log(1 - v))^{\frac{1}{\theta}} + (\phi_2 \log(u))^{\frac{1}{\theta}} \right]^\theta. \quad (1.37)$$

1.1.5 Copules et formes de dépendance

1.1.5.1 La copule d'indépendance

Dans le cas où les variables X et Y sont indépendantes, nous avons

Théorème 1.3. *Soient X et Y deux variables aléatoires continues. Alors X et Y sont indépendantes si et seulement si*

$$C_{XY} = \pi, \quad (1.38)$$

où $\pi(u, v) = uv$. π est alors appelée copule indépendance ou copule produit.

Démonstration. La preuve de ce théorème découle du théorème 1.1 et le fait que X et Y soient indépendantes si et seulement si $H(x, y) = F(x) \times G(y)$ pour tout $x, y \in \bar{R}^2$. \square

Remarque 1.3. *Deux autres cas particuliers de copules existent, ce sont les cas où C est soit*

$$W(u, v) = \max(0, u + v - 1) \quad (1.39)$$

$$\text{ou} \quad M(u, v) = \min(u, v), \quad (1.40)$$

appelées les limites de Fréchet-Hoeffding qui sont elles-mêmes des copules. Nous pouvons aisément vérifier les conditions de la définition 1.3 pour chacune d'elles, (u, v) étant dans $[0, 1]^2$. En général, la copule C représente un modèle de dépendance si

$$W(u, v) \leq C(u, v) \leq M(u, v) \quad (1.41)$$

pour $(u, v) \in [0, 1]^2$, $\forall C$ copule (voir Genest and Favre [2007]).

1.1.5.2 Relation entre les copules et les mesures de dépendance

Il existe plusieurs mesures de la force et de la direction de la dépendance entre deux variables aléatoires. Les plus courantes sont la corrélation de Pearson, le tau de Kendall et le rho de Spearman. Le tau de Kendall et le rho de Spearman peuvent être exprimés en termes de copule correspondante, alors que ce n'est pas toujours le cas pour la corrélation de Pearson Czado [2019]. Ci-après nous

rappelons les expressions du tau de Kendall et du rho de Spearman ainsi que les formules qui les relient aux copules.

Rappelons d'abord la définition des notions de concordance et de discordance.

Définition 1.9. (Czado [2019]) La paire (x_i, x_j) est dite :

- concordante si l'ordre dans $x^1 := (x_{i1}, x_{j1})$ est le même que dans $x^2 := (x_{i2}, x_{j2})$, c.-à-d., soit nous avons $x_{i1} < x_{j1}$ et $x_{i2} < x_{j2}$ ou bien $x_{i1} > x_{j1}$ et $x_{i2} > x_{j2}$,
- discordante si l'ordre dans x^1 est opposé à l'ordre de x^2 , c.-à-d., soit nous avons $x_{i1} < x_{j1}$ et $x_{i2} > x_{j2}$ ou bien $x_{i1} > x_{j1}$ et $x_{i2} < x_{j2}$.

1.1.5.2.1 Le Tau de Kendall Le Tau de Kendall (τ) est un coefficient de corrélation très utilisé du fait qu'il soit robuste aux valeurs aberrantes et aux petits échantillons, qu'il ne requiert pas de relation linéaire ni de distribution normale (voir *Hennemuth et al.* [2013]). Il est donné sous la forme,

$$\tau = \frac{C - D}{n(n - 1)^2}, \quad (1.42)$$

C et D étant respectivement les nombres de paires de données concordantes et discordantes dans un certain échantillon aléatoire.

Le τ de Kendall joue un rôle très important dans la théorie des copules vu qu'il peut être exprimé en fonction des copules et vice-versa. Le théorème qui régit cette relation est comme suit

Théorème 1.4. (Nelsen [2007]; Czado [2019]) Soient (X_1, Y_1) et (X_2, Y_2) des vecteurs indépendants de variables aléatoires continues de fonctions de distributions jointes H_1 et H_2 et de fonctions marginales communes F pour X_1 et X_2 et G pour Y_1 et Y_2 . Soient C_1 et C_2 les copules respectives de (X_1, Y_1) et (X_2, Y_2) et soit K la différence entre les probabilités de concordance et discordance de (X_1, Y_1) et (X_2, Y_2)

$$K = P[(X_1 - X_2)(Y_1 - Y_2) > 0] - P[(X_1 - X_2)(Y_1 - Y_2) < 0], \quad (1.43)$$

alors le τ de Kendall a l'expression suivante,

$$\tau_{X,Y} = \tau_C = K(C, C) = 4 \iint_{I^2} C(u, v) dC(u, v) - 1 \quad (1.44)$$

Nous reprenons ici la démonstration de ce théorème donnée dans *Nelsen* [2007].

Démonstration. Les variables aléatoires étant continues, nous pouvons écrire

$$\begin{aligned}
 & P[(X_1 - X_2)(Y_1 - Y_2) < 0] = 1 - P[(X_1 - X_2)(Y_1 - Y_2) > 0], \\
 \text{d'où} & \quad \quad \quad K = 2P[(X_1 - X_2)(Y_1 - Y_2) > 0] - 1, \\
 \text{avec} & \quad P[(X_1 - X_2)(Y_1 - Y_2) > 0] = P[X_1 > X_2, Y_1 > Y_2] + P[X_1 < X_2, Y_1 < Y_2],
 \end{aligned}$$

ces probabilités sont équivalentes à

$$\begin{aligned}
 P[X_1 > X_2, Y_1 > Y_2] &= P[X_2 < X_1, Y_2 < Y_1] \\
 &= \iint_{\mathbb{R}^2} P[X_2 \leq x, Y_2 \leq y] dC_1(F(x), G(y)) \\
 &= \iint_{\mathbb{R}^2} C_2(F(x), G(y)) dC_1(F(x), G(y)),
 \end{aligned}$$

en utilisant $u = F(x)$ et $v = G(y)$ nous obtenons

$$P[X_1 > X_2, Y_1 > Y_2] = \iint_{\mathbb{I}^2} C_2(u, v) dC_1(u, v).$$

De façon similaire, nous avons

$$\begin{aligned}
 P[X_1 < X_2, Y_1 < Y_2] &= \iint_{\mathbb{R}^2} P[X_2 > x, Y_2 > y] dC_1(F(x), G(y)) \\
 &= \iint_{\mathbb{R}^2} P[1 - F(x) - G(y) + C_2(F(x), G(y))] dC_1(F(x), G(y)) \\
 &= \iint_{\mathbb{I}^2} [1 - u - v + C_2(u, v)] dC_1(u, v).
 \end{aligned}$$

C_1 étant la distribution jointe de la paire $(U, V) \sim U([0, 1])$ alors $E(U) = E(V) = \frac{1}{2}$, d'où

$$\begin{aligned}
 P[X_1 < X_2, Y_1 < Y_2] &= 1 - \frac{1}{2} - \frac{1}{2} + \iint_{\mathbb{I}^2} C_2(u, v) dC_1(u, v) \\
 &= \iint_{\mathbb{I}^2} C_2(u, v) dC_1(u, v), \\
 \text{alors} \quad P[(X_1 - X_2)(Y_1 - Y_2) > 0] &= 2 \iint_{\mathbb{I}^2} C_2(u, v) dC_1(u, v).
 \end{aligned}$$

Finalement, en remplaçant dans 1.43, nous obtenons l'équation 1.44 recherchée. □

Pour les copules Archimédiennes, *Nelsen* [2007] a montré que la forme de τ_C devient

$$\tau_C = 1 + 4 \int_0^1 \frac{\phi(t)}{\phi'(t)} dt \tag{1.45}$$

où ϕ est la fonction génératrice de la copule Archimédienne.

Nous donnons ci-après deux exemples d'obtention du τ de Kendall à partir des fonctions génératrices des copules Archimédiennes Clayton et Gumbel.

Exemple 1.12. Pour la copule de Clayton, dont la dérivée de la fonction génératrice est $\phi'_C(t) = -t^{-\theta-1}$ où $\theta > 0$, nous obtenons

$$\begin{aligned}\tau_C &= 1 + 4 \int_0^1 \frac{1-t^{-\theta}}{\theta t^{-\theta-1}} dt \\ &= 1 + \frac{4}{\theta} \int_0^1 -t + t^{\theta+1} dt \\ &= \frac{\theta}{\theta+2}\end{aligned}\tag{1.46}$$

Exemple 1.13. Pour la copule de Gumbel, dont la dérivée de la fonction génératrice est $\phi'_G(t) = \frac{-\theta(-\ln t)^{\theta-1}}{t}$ où $\theta \geq 1$, nous obtenons

$$\begin{aligned}\tau_C &= 1 + 4 \int_0^1 \frac{-t(-\ln t)^\theta}{\theta(-\ln t)^{\theta-1}} dt \\ &= 1 + \frac{4}{\theta} \int_0^1 t \ln t dt \\ &= 1 - \frac{1}{\theta} \quad \text{en intégrant par parties.}\end{aligned}\tag{1.47}$$

Nous reprenons également les expressions du τ de Kendall dans le cas des copules de Frank et Joe telles que données dans *Czado* [2019].

Exemple 1.14. L'expression du τ de Kendall en fonction du paramètre $\delta \in [-1, 1]$ de la copule de Frank est :

$$\tau = 1 - \frac{4}{\delta} + 4 \frac{D_1(\delta)}{\delta}, \quad \text{avec } D_1(\delta) = \int_0^\delta \frac{x/\delta}{e^x - 1} dx \text{ (fonction de Debye).}\tag{1.48}$$

Exemple 1.15. L'expression du τ de Kendall en fonction du paramètre $\delta \in [0, 1]$ de la copule de Joe est :

$$\tau = 1 + \left(\frac{-2 + 2\gamma + 2\ln(2) + \psi\left(\frac{1}{\delta}\right) + \psi\left(\frac{1}{2} \frac{2+\delta}{\delta}\right) + \delta}{-2 + \delta} \right),\tag{1.49}$$

avec $\gamma = \lim_{n \rightarrow \infty} \left(\sum_{i=1}^n \frac{1}{i} - \ln(n) \right) \approx 0,57721$ la constante d'Euler et $\psi(x) = \frac{d}{dx} \ln(\Gamma(x)) = \frac{d}{dx} \frac{\Gamma'(x)}{\Gamma(x)}$ la fonction digamma.

1.1.5.2.2 Le rho de Spearman

Définition 1.10. *Czado* [2019] Pour des variables aléatoires continues X_1 et X_2 de distributions marginales F_1 et F_2 , respectivement, le ρ_s de Spearman ou la corrélation de rang est défini comme la corrélation de Pearson des variables aléatoires $F_1(X_1)$ et $F_2(X_2)$, c.-à-d.,

$$\rho_s := \rho_s(X_1, X_2) = \text{Cor}(F_1(X_1), F_2(X_2)). \quad (1.50)$$

La relation entre le ρ_s de Spearman et la copule est donnée par le théorème suivant :

Théorème 1.5. *Nelsen* [2003, 2007] Soient X et Y des variables aléatoires continues de copule C . Alors la version de population du rho de Spearman pour X et Y (dénotée par ρ_{XY} ou ρ_C) est donnée par

$$\rho_{XY} = \rho_C = \int \int_{[0,1]^2} C(u, v) dudv - 3. \quad (1.51)$$

Dans le paragraphe suivant, nous allons définir un concept très utilisé dans la théorie des valeurs extrêmes, la dépendance dans les queues de distribution d'une paire de variables aléatoires.

1.1.5.3 Dépendance des queues de distributions des copules

De nombreux concepts de dépendance sont conçus pour décrire comment les différentes valeurs d'une variable aléatoire évoluent avec celles d'une autre. Un concept de ce type est la dépendance de queue, qui mesure la dépendance entre les variables dans le quadrant supérieur droit et le quadrant inférieur gauche de \mathbb{I}^2 (voir *Nelsen* [2007]). Pour évaluer la dépendance de la queue, nous considérons la probabilité d'apparition jointe de valeurs extrêmement faibles ou grandes (voir *Czado* [2019]).

Définition 1.11. (Coefficients de dépendance des queues supérieure et inférieure) Le coefficient de dépendance de la queue supérieure d'une distribution bivariable de copule C est défini comme,

$$\begin{aligned} \lambda^u &:= \lambda^{upper} = \lim_{u \rightarrow 1^-} P(X_2 > F_2^{-1}(u) / X_1 > F_1^{-1}(u)) \\ &= \lim_{u \rightarrow 1^-} \frac{1 - 2u + C(u, u)}{1 - u}. \end{aligned} \quad (1.52)$$

De façon similaire, le coefficient de dépendance de la queue inférieure est donné par,

$$\begin{aligned} \lambda^l &:= \lambda^{lower} = \lim_{u \rightarrow 0^+} P(X_2 \leq F_2^{-1}(u) / X_1 \leq F_1^{-1}(u)) \\ &= \lim_{u \rightarrow 0^+} \frac{C(u, u)}{u}. \end{aligned} \quad (1.53)$$

Nous donnons ci-après quelques résultats de calcul des coefficients de dépendance des queues des copules bivariées utilisées dans cette étude en fonction de leurs paramètres, tels que présentés dans *Czado* [2019].

Pour les copules de Gumbel et de Joe, nous avons $\lambda^u = 2 - 2^{1/\delta}$ et $\lambda^l = 0$, elles sont donc adéquates pour la représentation des dépendances dans les queues supérieures mais pas de celles dans les queues inférieures, contrairement à la copule de Clayton pour laquelle nous avons $\lambda^u = 0$ et $\lambda^l = 2^{1/\delta}$. Pour la copule de Frank, nous avons $\lambda^u = \lambda^l = 0$, son utilisation n'est donc pas souhaitable pour représenter les distributions jointes avec des dépendances dans les queues. Pour la copule de Tawn, nous avons $\lambda^u = (\psi_1 + \psi_2) - (\psi_1^\theta + \psi_2^\theta)^{1/\theta}$ et $\lambda^l = 0$, elle peut donc être utilisée de la même sorte que les copules de Gumbel et de Joe. Ceci dit, la rotation des copules offre la possibilité de représenter les autres types de dépendance comme discuté dans la sous-section 1.1.4.

1.2 Construction et estimation des copules

1.2.1 Algorithme général

Comme nous l'avons vu dans la section précédente, le théorème de Sklar assure l'existence de la copule entre deux (ou plusieurs) variables aléatoires continues à fonctions de distributions univariées. Cependant, le théorème en soi ne permet pas de désigner une copule particulière pour illustrer la distribution jointe. Dans ce qui suit, nous expliquons les étapes d'obtention des copules à partir des marges univariées en utilisant des méthodes graphiques et analytiques.

Soit un échantillon $(X_1, Y_1) \dots (X_n, Y_n)$ du couple de variables aléatoires (X, Y) , l'algorithme suivant permet d'identifier les copules candidates pour modéliser la distribution jointe de (X, Y) :

Étape 1 Utilisation d'une approche graphique sous forme de K-plots ou de Chi-plots pour avoir une première idée sur les familles de copules candidates ajustées aux données.

Étape 2 Utilisation du Critère d'Information d'Akaike (AIC) pour réduire l'ensemble de candidats obtenu à l'étape 1.

Étape 3 Utilisation de tests d'ajustement pour sélectionner les meilleurs candidats.

Dans le paragraphe suivant, nous décrivons l'étape préliminaire à l'utilisation de cet algorithme, celle où les distributions marginales sont construites.

1.2.2 Modélisation des marges

Avant d'expliquer comment obtenir les marges à partir des données initiales, nous jugeons utile d'introduire les définitions suivantes :

Une série chronologique est une succession d'observations au cours du temps $\{X_t : t = 1, \dots, n\} = \{X_1, \dots, X_n\}$ dont l'analyse a pour objectif de caractériser la dépendance qui peut exister dans une série d'observations et peut servir à donner des prévisions de la série.

La stationnarité est l'une des propriétés les plus importantes des séries chronologiques. Elle est généralement obtenue en supprimant la tendance et la saisonnalité dans la série. Elle est définie par :

Définition 1.12. Une suite $\{X_t : t = 1, \dots, n\}$ de variables aléatoires est dite stationnaire du second ordre si elle vérifie les propriétés suivantes :

$$\mathbb{E}[X_t] = \mu < \infty; \quad (1.54)$$

$$\mathbb{E}[X_t^2] < \infty; \quad (1.55)$$

$$\text{Cov}(X_s, X_{s+t}) < \text{Cov}(X_{s-1}, X_{s-1+t}) = \dots = \text{Cov}(X_1, X_t), \text{ pour tout } s, t \in \mathbb{N}, \quad (1.56)$$

où $\text{Cov}(X_t, X_{t+h})$ est l'autocovariance à l'horizon h notée $\gamma^X(h)$. Cette dernière, lorsqu'elle est divisée par la variance de X_t donne la fonction d'autocorrélation (ACF), notée $\rho^X(h) = \frac{\gamma^X(h)}{\gamma^X(0)}$ qui a pour propriétés :

1. $-1 \leq \rho^X(h) \leq 1$;
2. $\rho^X(h) = 0$ signifie que les observations X_t et $X_t + h$ sont non corrélées ;
3. $\rho^X(h) = 1$ ou $\rho^X(h) = -1$ illustre une corrélation parfaite.

Rappelons aussi la définition d'un processus Autorégressif Moyenne Mobile (ARMA) :

Définition 1.13 (*Lagnoux [2015a]*). On appelle **processus autorégressif moyenne mobile** d'ordre (p, q) tout processus $(X_t)_{t \in \mathbb{Z}}$ stationnaire tel que

$$\forall t \in \mathbb{Z}, X_t = \varepsilon_t + \sum_{i=1}^p \phi_i X_{t-i} + \sum_{i=1}^q \theta_i \varepsilon_{t-i}, \quad (1.57)$$

où $\phi_1, \phi_2, \dots, \phi_p, \theta_1, \theta_2, \dots, \theta_q$ sont des réels fixés et $(\varepsilon_t)_{t \in \mathbb{Z}}$ est un bruit blanc de variance σ_ε^2 . Un tel processus est dit *ARMA*(p, q) (Autoregressive Moving Average of order (p, q)).

Le traitement des données observées $((X_1, \dots, X_T)$ ou $(Y_1, \dots, Y_T))$ diffère selon qu'elles présentent des autocorrélations ou non :

1. Dans le premier cas, des modèles de séries chronologiques appropriés sont ajustés à chaque série de variables. Concrètement, il s'agit de modèles ARMA qui sont ajustés (en utilisant la méthode générale de *Box and Jenkins [1970]*¹) à chacune des deux marges pour tenir compte des dépendances temporelles marginales. Dans un second temps, les résidus normalisés sont formés. Ces derniers étant des variables aléatoires indépendantes et identiquement distribuées (i.i.d)², ils peuvent être utilisés comme données pour les copules après

1. Ces étapes sont par ordre : le choix des ordres AR et MA ; l'estimation des paramètres ; le calcul des résidus ; les tests sur les résidus ; l'itération des étapes précédentes si les résidus ne sont pas des bruits blancs. Elles sont décrites en détail dans *Lagnoux [2015b]* et dans plusieurs autres ouvrages sur les séries chronologiques.

2. Pour satisfaire la condition i.i.d. des résidus, nous effectuons des tests d'indépendance (Box-Pierce et Ljung-Box) et des tests de normalité (Anderson-Darling et Kolmogorov-Smirnov). Pour les rappels théoriques concernant ces tests voir Annexe A.1, page 87.

une transformation dans $[0, 1]$ (voir *Czado* [2019]). La dépendance entre ces résidus peut donc être modélisée par des copules bivariées.

2. Dans le deuxième cas, le théorème de Sklar permet la modélisation directe des dépendances grâce aux copules conditionnelles.

Plus de détails concernant cette approche sont donnés dans *Okou* [2014].

Dans le paragraphe suivant, nous décrivons comment l'algorithme introduit précédemment s'applique lorsque les observations sont des séries chronologiques.

Soit $(X_t)_{t \geq 0}$ une série chronologique, X_{t_1}, \dots, X_{t_n} un ensemble fini d'observations, et $\varepsilon_{t_1}, \dots, \varepsilon_{t_n}$ les résidus correspondants. Soit $U_i = F(\varepsilon_i)$, où F est la CDF de ε_i . On sait que les U_i sont uniformément distribués sur $[0, 1]$. Ordonnons les résidus et notons par r_{t_j} le rang de ε_{t_j} , et soit $U_{t_j} = \frac{r_{t_j}}{n+1}$. Ce dernier est une estimation de $F(\varepsilon_{t_j})$. Soit $(Y_t)_{t \geq 0}$ une autre série chronologique à partir de laquelle nous obtenons le vecteur $(V_{t_1}, \dots, V_{t_k})$ de la même manière que pour $(U_{t_1}, \dots, U_{t_k})$. Nous utilisons les paires (U_{t_j}, V_{t_j}) pour obtenir un K-plot pour évaluer la dépendance entre les deux séries temporelles $(X_t)_t$ et $(Y_t)_t$.

La construction des Kendall-plots constitue la première étape de l'algorithme décrit dans le paragraphe 1.2.1, page 24 nous en donnons ci-après quelques détails.

1.2.3 Les Kendall-plots

Pour les données continues multivariées, l'indépendance stochastique est une caractéristique de la copule sous-jacente et, par conséquent, les tests d'association graphiques et formels doivent être basés sur les rangs, car ces derniers sont les statistiques les plus invariantes sous des transformations qui laissent la copule inchangée (voir *Genest and Boies* [2003]). Les graphiques qu'on utilise pour évaluer la dépendance dans un échantillon bivarié de variables aléatoires, appelés K-plots, répondent à cette exigence de base et bénéficient d'autres propriétés qui en font des outils complémentaires pour l'inspection visuelle des données (voir *Genest and Boies* [2003]).

La procédure des K-plots consiste à représenter les paires $(W_{i,n}, H_i)$ pour $i \in [1, n]$. Pour une

paire de variables aléatoires donnée (X_i, Y_i) avec $1 \leq i \leq n$, soit

$$H_i = \frac{1}{n-1} \#\{j \neq i : X_j \leq X_i, Y_j \leq Y_i\} \quad (1.58)$$

et

$$W_{i:n} = n \binom{n-1}{i-1} \int_0^1 w \{K_0(w)\}^{i-1} \times \{1 - K_0(w)\}^{n-i} dK_0(w), \quad (1.59)$$

où $W_{i:n}$ représente l'espérance de la statistique d'ordre i dans un échantillon de taille n de la distribution K_0 des H_i sous l'hypothèse nulle d'indépendance de X et Y avec

$$K(w) = Pr\{H(X, Y) \leq w\}, 0 \leq w \leq 1. \quad (1.60)$$

Ensuite, il suffit d'évaluer l'éloignement des points $(W_{i:n}, H_i)$ du K-plot par rapport à la ligne diagonale pour se renseigner sur le degré de dépendance entre X et Y .

Les K-plots aident aussi à déterminer la copule caractérisant une certaine structure de dépendance, grâce à la courbe qu'ils affichent. De plus, les K-plots ont l'avantage d'être facilement extensibles au contexte multivarié (voir *Genest and Boies* [2003]). Dans *Genest and Boies* [2003], nous pouvons trouver un catalogue intéressant des cas de figures des K-plots suivant les différents degrés de dépendance possibles. Les interprétations données à chaque cas constituent une référence dans la matière. Nous trouvons par exemple dans *Ait Hassou et al.* [2017] comment peuvent être utilisés les K-plots dans la sélection des copules candidates.

Ceci dit, le choix des familles de copules basé sur des outils graphiques reste largement subjectif et nécessite une certaine expérience. Aussi, nous poursuivons notre analyse en utilisant quelques outils analytiques pour choisir les copules candidates les plus appropriées. Notre liste initiale des copules se limite à celles les plus fréquemment utilisées en hydrologie (voir *Golian et al.* [2012]), à savoir les copules Archimédiennes de Clayton, Frank, Gumbel et Joe auxquelles nous ajouterons la copule de Tawn pour tenir en compte des éventuelles représentations de valeurs extrêmes. Mais tout d'abord, nous devons effectuer un test d'indépendance.

1.2.4 Test d'indépendance

Nous allons inclure la copule d'indépendance dans la sélection en effectuant un test d'indépendance préliminaire basé sur le tau de Kendall comme décrit dans *Genest and Favre* [2007] et *Dißmann et al.* [2013]. Si ce test indique l'indépendance, aucune autre mesure n'est prise et la copule d'indépendance est choisie. Ce test exploite la normalité asymptotique de la statistique de test

T donnée par

$$T = \left(\frac{9N(N-1)}{2(2N+5)} \right)^{0.5} \times |\hat{\tau}|, \quad (1.61)$$

où N est le nombre d'observations (longueur de u_1) et $\hat{\tau}$ le tau de Kendall empirique des vecteurs de données u_1 et u_2 . La valeur p de l'hypothèse nulle d'indépendance bivariée est donc asymptotiquement

$$p = 2 \times (1 - \phi(T)), \quad (1.62)$$

où ϕ est la CDF normale standard.

La suite des étapes concernera donc uniquement les paires de variables pour lesquelles les résultats du test rejettent l'hypothèse d'indépendance.

1.2.5 Estimation des paramètres

Il existe plusieurs méthodes pour l'estimation des paramètres d'une copule (voir *Nelsen* [2007] et *Joe* [2014]). Les plus connues, telles que décrites dans *Singh and Zhang* [2018], sont l'Estimation par la méthode du Maximum de Vraisemblance (MLE) (Full-MLE) où les paramètres des distributions marginales et des fonctions de copule sont estimés simultanément. La MLE est appliquée en deux étapes où on estime d'abord les paramètres des distributions marginales à partir desquels les paramètres de la fonction de copule sont estimés. La MLE semi-paramétrique (ou pseudo MLE) où les paramètres de la fonction de copule sont estimés à partir des marges empiriques.

Parmi les méthodes pré-décrites, nous utilisons la méthode pseudo MLE. Nous avons donc moins de chance de commettre des erreurs d'estimation liées à l'identification des distributions marginales (voir *Singh and Zhang* [2018]).

1.2.6 Comparaison des AIC

Afin de sélectionner les copules paramétriques bivariées, nous utilisons le critère Akaike information criterion (AIC) (1998) pour choisir entre les familles de copules bivariées avec leurs paramètres. Cette approche a été suivie par *Manner* [2007] et *Brechmann* [2010]. Particulièrement, la sélection de copule bivariée basée sur l'AIC a été comparée dans une étude basée sur la méthode de Monte Carlo à grande échelle (voir *Brechmann* [2010]) à plusieurs autres alternatives telles que la sélection de la famille avec la plus grande p -valeur d'un test d'adéquation de la copule basé sur la statistique de Cramér-von Mises (voir *Genest et al.* [2009]) et la plus petite distance entre les caractéristiques de dépendance empiriques et modélisées (τ de Kendall et dépendance dans les queues

des distributions). L'AIC s'est avéré être le critère de sélection le plus fiable pour les copules bivariées (voir aussi *Czado* [2019]).

Pour chaque copule candidate, le calcul de l'AIC se fait comme suit (voir *Manner* [2007] et *Brechmann* [2010]),

$$AIC := -2\sum_{i=1}^N \ln[c(u_{i,1}, u_{i,2} | \theta)] + 2k, \quad (1.63)$$

où k est le nombre de paramètres θ de la copule C .

Bien que l'AIC soit un bon critère pour trouver la copule la mieux adaptée, ses performances ne sont pas entièrement satisfaisantes dans le cas de faibles dépendances (voir *Manner* [2010]). En effet, la sélection des modèles doit refléter l'adéquation des caractéristiques des copules avec les différentes propriétés des données. Une certaine réflexion est donc recommandée lors de l'utilisation de l'AIC, même si, dans la plupart des cas, lorsque le vrai modèle n'est pas sélectionné, le modèle avec l'AIC le plus bas est un modèle avec une structure de dépendance similaire (voir *Manner* [2010]).

1.2.7 Test d'ajustement

Pour les couples d'observations non indépendants, nous effectuons le test de qualité d'ajustement des copules de Cramér-von Mises (voir *Genest et al.* [2009]) basé sur la copule empirique (voir l'équation 1.3, page 7). Dans *Genest et al.* [2009], des simulations basées sur des procédures de bootstrap paramétriques sont utilisées pour approximer la distribution nulle des statistiques des tests de Cramér-von Mises et de Kolmogorov-Smirnov. En comparant ces statistiques relativement à un même processus empirique, la statistique de Cramér-von Mises est presque invariablement la plus puissante.

La statistique du test de Cramér-von Mises est donnée par

$$S_n = \int_{[0,1]^2} \mathbb{C}_n(u)^2 dC_n(u). \quad (1.64)$$

où $\mathbb{C}_n = \sqrt{n}(C_n - C_{\theta_n})$. De grandes valeurs de cette statistique mènent au rejet de l'hypothèse

$$H_0 : C \in C_0, C_0 = \{C_\theta : \theta \in \mathcal{O}\}, \quad (1.65)$$

avec \mathcal{O} un sous ensemble ouvert de \mathbb{R}^p , pour $p \geq 1$. L'approximation de la p-value de cette statis-

tique est

$$p = 0,5 + \sum_{b=1}^N \frac{T[b] \geq T}{N+1}, \quad (1.66)$$

où T est la statistique test, et $T[b]$ la statistique test obtenue par la technique de Bootstrap.

1.2.8 Préviation

Après avoir identifié la bonne copule de la paire de variables aléatoires (U, V) , nous utilisons sa forme conditionnelle pour obtenir les valeurs v de V en utilisant les observations u de U . La copule conditionnelle proposée par *Nelsen* [2007] est de la forme

$$\begin{aligned} C_u(v) &= P[V \leq v | U = u] \\ &= \lim_{\Delta u \rightarrow 0} \frac{C(u + \Delta u, v) - C(u, v)}{\Delta u} \\ &= \frac{\partial C(u, v)}{\partial u}. \end{aligned} \quad (1.67)$$

De la même façon, nous pouvons obtenir les valeurs u en utilisant les observations v :

$$\begin{aligned} C_v(u) &= P[U \leq u | V = v] \\ &= \lim_{\Delta v \rightarrow 0} \frac{C(u, v + \Delta v) - C(u, v)}{\Delta v} \\ &= \frac{\partial C(u, v)}{\partial v}. \end{aligned} \quad (1.68)$$

Exemple 1.16. A partir des équations 1.26 et 1.67, nous obtenons pour la copule de survie de Joe :

$$\begin{aligned} \hat{C}_{J_u}(v) &= \frac{\partial \left(u + v - (u^\theta + v^\theta - u^\theta v^\theta)^{\frac{1}{\theta}} \right)}{\partial u} \\ &= 1 - \frac{\partial \left((u^\theta + v^\theta - u^\theta v^\theta)^{\frac{1}{\theta}} \right)}{\partial u} \\ &= 1 - \frac{1}{\theta} \left(\theta u^{(\theta-1)} - \theta u^{(\theta-1)} v^\theta \right) \left(u^\theta + v^\theta - u^\theta v^\theta \right)^{\left(\frac{1}{\theta}-1\right)} \\ &= 1 - u^{(\theta-1)} (1 - v^\theta) \left(u^\theta (1 - v^\theta) + v^\theta \right)^{\left(\frac{1}{\theta}-1\right)} \end{aligned} \quad (1.69)$$

Exemple 1.17. A partir des équations 1.18 et 1.67, nous obtenons pour la copule Tawn Type 1 :

$$\begin{aligned}
C_{T1u}(v) &= \frac{\partial \left(uv - v^{(\theta+\phi)} \right) \times \exp \left[\frac{(\log(v))^2}{\log(uv)} \left(\theta + \phi \frac{\log(v)}{\log(uv)} \right) \right]}{\partial u} \\
&= v \exp \left[\frac{(\log(v))^2}{\log(uv)} \left(\theta + \phi \frac{\log(v)}{\log(uv)} \right) \right] + \left(uv - v^{(\theta+\phi)} \right) \frac{\partial \exp \left[\frac{(\log(v))^2}{\log(uv)} \left(\theta + \phi \frac{\log(v)}{\log(uv)} \right) \right]}{\partial u} \\
&= \exp \left[\frac{(\log(v))^2}{\log(uv)} \left(\theta + \phi \frac{\log(v)}{\log(uv)} \right) \right] \left[v - \frac{(\log v)^2}{u} \left(uv - v^{(\theta+\phi)} \right) \left(\theta + 2\phi \frac{\log v}{\log(uv)} \right) \right]
\end{aligned} \tag{1.70}$$

Exemple 1.18. A partir des équations 1.35 et 1.67, nous obtenons pour la copule Tawn Type 2 :

$$\begin{aligned}
\hat{C}_{T2u}(v) &= \frac{\partial \left(u + v - 1 + (1-u)^{(1-\phi_1)}(1-v)^{(1-\phi_2)} \exp \left[(\phi_1 \log(1-v))^{\frac{1}{\theta}} + (\phi_2 \log(1-u))^{\frac{1}{\theta}} \right]^\theta \right)}{\partial u} \\
&= 1 + (1-v)^{(1-\phi_2)} \left(\frac{\partial \left((1-u)^{(1-\phi_1)} \right)}{\partial u} \exp \left[(\phi_1 \log(1-v))^{\frac{1}{\theta}} + (\phi_2 \log(1-u))^{\frac{1}{\theta}} \right]^\theta + \right. \\
&\quad \left. (1-u)^{(1-\phi_1)} \frac{\partial \left(\exp \left[(\phi_1 \log(1-v))^{\frac{1}{\theta}} + (\phi_2 \log(1-u))^{\frac{1}{\theta}} \right]^\theta \right)}{\partial u} \right)
\end{aligned} \tag{1.71}$$

avec

$$\begin{aligned}
\frac{\partial \left(\exp \left[(\phi_1 \log(1-v))^{\frac{1}{\theta}} + (\phi_2 \log(1-u))^{\frac{1}{\theta}} \right]^\theta \right)}{\partial u} &= \exp \left[(\phi_1 \log(1-v))^{\frac{1}{\theta}} + (\phi_2 \log(1-u))^{\frac{1}{\theta}} \right]^\theta \times \\
&\quad \frac{\phi_2}{u-1} \left[(\phi_1 \log(1-v))^{\frac{1}{\theta}} + (\phi_2 \log(1-u))^{\frac{1}{\theta}} \right]^{\theta-1}
\end{aligned} \tag{1.72}$$

en remplaçant 1.72 dans 1.71, nous obtenons

$$\begin{aligned}
\hat{C}_{T2u}(v) &= 1 + (1-v)^{(1-\phi_2)} \exp \left[(\phi_1 \log(1-v))^{\frac{1}{\theta}} + (\phi_2 \log(1-u))^{\frac{1}{\theta}} \right]^\theta \times \left((\phi_1 - 1)(1-u)^{-\phi_1} + \right. \\
&\quad \left. \frac{\phi_2}{u-1} \left[(\phi_1 \log(1-v))^{\frac{1}{\theta}} + (\phi_2 \log(1-u))^{\frac{1}{\theta}} \right]^{\theta-1} \right)
\end{aligned} \tag{1.73}$$

1.3 Les copules-vignes

Dans cette section, nous allons rappeler les définitions des notions de vigne, de R-vigne et R-copule-vigne qui constitueront notre environnement de travail.

1.3.1 La R-copule-vigne

Commençons par rappeler quelques définitions élémentaires à la spécification de la R-copule-vigne correspondant à une R-vigne telles que définies dans *Bedford and Cooke* [2002].

Définition 1.14. V est une vigne sur n éléments (variables) si :

1. $V = (T_1, \dots, T_m)$, pour $m \leq n$.
2. T_1 est un arbre de noeuds $N_1 = \{1, \dots, n\}$ et un ensemble d'arêtes noté E_1 .
3. Pour $i = 2, \dots, m$, T_i est un arbre de noeuds $N_i \subset N_1 \cup E_1 \cup E_2 \cup \dots \cup E_{i-1}$ et un ensemble d'arêtes noté E_i .

V est dite une *Vigne régulière* quand $m = n$, T_i est un arbre connecté à l'ensemble des arêtes E_i et à l'ensemble des noeuds $N_i = E_{i-1}$, de cardinal $\#N_i = n - (i - 1)$ pour $i = 1, \dots, n$. En plus, la condition de proximité doit être vérifiée, c-à-d pour $i = 2, \dots, n - 1$, si $a = \{a_1, a_2\}$ et $b = \{b_1, b_2\}$ sont deux noeuds dans N_i connectés par un arête $(a_1, a_2, b_1$ et $b_2)$, alors $\#a \cap b = 1$.

Définition 1.15. (spécification de la R-vigne). (F, V, B) est une spécification de la R-vigne si $F = (F_1, \dots, F_n)$ est un vecteur de fonctions de distributions marginales continues, V est une R-vigne à n dimensions et pour un ensemble de copules bivariées B_e nous avons

$$B = B_e / i = 1, \dots, n - 1; e \in E_i. \quad (1.74)$$

En notant pour tout $e_i \in E_i$, l'*union complète* de e_i le sous-ensemble

$$A_{e_i} = \{j \in N_1 = E_0 / \exists 1 \leq i_1 \leq i_2 \leq \dots \leq i_r = i \text{ et } e_{i_k} \in E_{i_k} (k = 1, \dots, r), \text{ avec } j \in e_{i_1}, e_{i_k} \in e_{i_{k+1}} (k = 1, \dots, r - 1)\}, \quad (1.75)$$

nous pouvons définir les ensembles suivants :

Définition 1.16. (Ensemble des contraintes). Pour $e = \{j, k\} \in E_i, i = 1, \dots, m - 1$, l'*ensemble conditionnant* associé à e est

$$D_e = A_j \cap A_k,$$

et les ensembles conditionnés associés à e sont

$$C_{e,j} = A_j - D_e \text{ et } C_{e,k} = A_k - D_e.$$

l'ensemble des contraintes pour V est

$$CV = \{(C_{e,j}, C_{e,k}, D_e) \mid i = 1, \dots, m-1, e \in E_i, e = \{j, k\}\}.$$

Notons que $A_e = A_j \cup A_k = C_{e,j} \cup C_{e,k} \cup D_e$ quand $e = \{j, k\}$. Pour $e \in E_m$, l'ensemble conditionnant est vide.

Bedford and Cooke [2001, 2002] ont montré que la densité d'une R-copule-vigne, spécifiée par l'attribution de copules bivariées appropriées aux arêtes de la R-vigne, est égale au produit des copules conditionnelles et inconditionnelles attribuées à ses bords.

Théorème 1.6. *Soit (F, V, B) une spécification R-vigne sur n éléments. Il existe une distribution de R-vigne copule unique F qui réalise cette spécification de densité*

$$f_{1, \dots, n}(x) = \prod_{k=1}^n f_k(x_k) \prod_{i=1}^{n-1} \prod_{e \in E_i} c_{C_{e,a}, C_{e,b}/D_e}(F_{C_{e,a}/D_e}(x_{C_{e,a}/x_{D_e}}), F_{C_{e,b}/D_e}(x_{C_{e,b}/x_{D_e}})) \quad (1.76)$$

où $x = (x_1, \dots, x_n)$, $e = a, b$ et x_{D_e} , représente les variables dans D_e , i.e., $x_{D_e} = x_i / i \in D_e$, f_i désigne la densité de F_i pour $i = 1, \dots, n$ et $c_{C_{e,a}, C_{e,b}/D_e}$ est la densité de copule de la paire $(C_{e,a}, C_{e,b})$ conditionnée par D_e .

1.3.2 Copules conditionnelles

La version conditionnelle du théorème de Sklar a été donnée par *Patton* [2001] :

Théorème 1.7. *Soit H une CDF conditionnelle bivariée avec des marges continues F et G , et soit \mathcal{F} un ensemble de conditionnement. Alors il existe une unique copule conditionnelle $C : [0, 1] \times [0, 1] \rightarrow [0, 1]$ telle que*

$$H(x, y / \mathcal{F}) = C(F(x / \mathcal{F}), G(y / \mathcal{F}) / \mathcal{F}), \forall x, y \in \mathbb{R} \quad (1.77)$$

Inversement, si C est une copule conditionnelle et que F et G sont les fonctions de distribution conditionnelles de deux variables aléatoires X et Y , alors la fonction H définie par l'équation 1.77 est une CDF conditionnelle bivariée de marges F et G .

Notons que lorsque F et G sont différentiables et H et C différentiables deux fois, la densité équivalente à l'équation 1.77, qui est très utile pour les analyses MLE, peut être obtenue comme

suit

$$\begin{aligned}
h(x, y / \mathcal{F}) &\equiv \frac{\partial^2 H(x, y / \mathcal{F})}{\partial x \partial y} \\
&= \frac{\partial^2 C(F(x / \mathcal{F}), G(y / \mathcal{F}) / \mathcal{F})}{\partial(F(x / \mathcal{F})) \partial(G(y / \mathcal{F}))} \cdot \frac{\partial(F(x / \mathcal{F}))}{\partial x} \cdot \frac{\partial(G(y / \mathcal{F}))}{\partial y} \\
&= \frac{\partial^2 C(u, v / \mathcal{F})}{\partial u \partial v} \cdot f(x / \mathcal{F}) \cdot G(y / \mathcal{F}) \\
&= c(u, v / \mathcal{F}) \cdot f(x / \mathcal{F}) \cdot G(y / \mathcal{F}), \forall x, y \in \bar{\mathbb{R}}
\end{aligned} \tag{1.78}$$

où $u \equiv F(x / \mathcal{F})$ et $v \equiv G(y / \mathcal{F})$.

Pour obtenir les copules conditionnelles dans le cas des R-copule-vignes, *Joe* [1996] a montré que :

$$\begin{aligned}
F_{C_{e,a}/D_e}(x_{C_{e,a}}/x_{D_e}) &= \frac{\partial C_{C_a/D_a}(F_{C_{a,a_1}/D_a}(x_{C_{a,a_1}}/x_{D_a}), F_{C_{a,a_2}/D_a}(x_{C_{a,a_2}}/x_{D_a}))}{\partial F_{C_{a,a_2}/D_a}(x_{C_{a,a_2}}/x_{D_a})} \\
&:= h(F_{C_{a,a_1}/D_a}(x_{C_{a,a_1}}/x_{D_a}), F_{C_{a,a_2}/D_a}(x_{C_{a,a_2}}/x_{D_a}))
\end{aligned} \tag{1.79}$$

où chaque terme de la dérivée partielle de la copule doit être obtenu récursivement.

Ci-après, nous donnons des exemples de calcul des fonctions h et h^{-1} pour les copules Clayton, Frank et Gumbel utilisées dans cette étude.

Exemple 1.19. Pour la copule de Clayton, l'équation 1.9, page 11 nous donne :

$$\begin{aligned}
\frac{\partial C_C(u, v)}{\partial u} &= \frac{-1}{\theta} (u^{-\theta} + v^{-\theta} - 1)^{\frac{-1}{\theta} - 1} \times \frac{\partial (u^{-\theta} + v^{-\theta} - 1)}{\partial u} \\
\text{d'où} \quad h_C(u, v) &= u^{-\theta-1} (u^{-\theta} + v^{-\theta} - 1)^{\frac{-1-\theta}{\theta}}
\end{aligned} \tag{1.80}$$

Pour calculer $h_C^{-1}(u, v)|_v$ posons $y = h_C(u, v) = \frac{\partial C_C(u, v)}{\partial u}$, nous avons donc :

$$\begin{aligned}
y^{\frac{-1}{1+\theta}} &= (1 + u^\theta (v^{-\theta} - 1))^{\frac{1}{\theta}} \\
\text{d'où} \quad y^{\frac{-\theta}{1+\theta}} &= 1 + u^\theta (v^{-\theta} - 1) \\
\text{ce qui est équivalent à} \quad u^\theta &= \frac{y^{\frac{-\theta}{1+\theta}} - 1}{v^{-\theta} - 1} \\
\text{Finalement, nous obtenons} \quad h_C^{-1}(u, v)|_v &= \left(\frac{u^{\frac{-\theta}{1+\theta}} - 1}{v^{-\theta} - 1} \right)^{\frac{1}{\theta}}
\end{aligned} \tag{1.81}$$

Exemple 1.20. Pour la copule de Frank, l'équation 1.10, page 12 nous donne :

$$\begin{aligned}
\frac{\partial C_F(u, v)}{\partial u} &= \frac{-1}{\theta} \frac{\partial \left[1 + \frac{(e^{-\theta u} - 1)(e^{-\theta v} - 1)}{e^{-\theta} - 1} \right]}{\partial u} \times \frac{1}{1 + \frac{(e^{-\theta u} - 1)(e^{-\theta v} - 1)}{e^{-\theta} - 1}} \\
&= \frac{-1}{\theta} \frac{1}{e^{-\theta} - 1} \frac{\partial [(e^{-\theta u} - 1)(e^{-\theta v} - 1)]}{\partial u} \times \frac{1}{1 + \frac{(e^{-\theta u} - 1)(e^{-\theta v} - 1)}{e^{-\theta} - 1}} \\
&= \frac{-1}{\theta} (-\theta) e^{-\theta u} \frac{e^{-\theta v} - 1}{e^{-\theta} - 1} \times \frac{1}{1 + \frac{(e^{-\theta u} - 1)(e^{-\theta v} - 1)}{e^{-\theta} - 1}}
\end{aligned}$$

d'où $h_F(u, v) = \frac{e^{-\theta u}(e^{-\theta v} - 1)}{(e^{-\theta} - 1) + (e^{-\theta u} - 1)(e^{-\theta v} - 1)}$ (1.82)

Posons $y = h_F(u, v)$, nous avons :

$$\begin{aligned}
y &= 1 + \frac{e^{-\theta v} - e^{-\theta}}{e^{-\theta u}(e^{-\theta v} - 1) + e^{-\theta} - e^{-\theta v}} \\
\text{d'où } e^{-\theta u}(e^{-\theta v} - 1) &= \frac{1 + e^{-\theta v} - e^{-\theta}}{y} - e^{-\theta} + e^{-\theta v} \\
\text{ce qui est équivalent à } e^{-\theta u} &= \frac{1 + (1 + y)(e^{-\theta v} - e^{-\theta})}{y(e^{-\theta v} - 1)}
\end{aligned}$$

Enfinement, nous obtenons $h_F^{-1}(u, v)|_v = \ln \left[\frac{u(e^{-\theta v} - 1)}{1 + (1 + u)(e^{-\theta v} - e^{-\theta})} \right]^{\frac{1}{\theta}}$ (1.83)

Exemple 1.21. Pour la copule de Gumbel, l'équation 1.11, page 12 nous donne :

$$\begin{aligned}
\frac{\partial C_G(u, v)}{\partial u} &= \frac{\partial \left(-[(-\ln(u))^\delta + (-\ln(v))^\delta] \right)}{\partial u} \times \exp -[(-\ln(u))^\delta + (-\ln(v))^\delta]^{\frac{1}{\delta}} \\
&= \frac{-1}{\delta} [(-\ln(u))^\delta + (-\ln(v))^\delta]^{\frac{1}{\delta} - 1} \times \frac{\partial}{\partial u} [(-\ln(u))^\delta + (-\ln(v))^\delta] \times \\
&\quad \exp -[(-\ln(u))^\delta + (-\ln(v))^\delta]^{\frac{1}{\delta}}
\end{aligned}$$

d'où $h_G(u, v) = \frac{(-\ln(u))^{\delta-1}}{u} [(-\ln(u))^\delta + (-\ln(v))^\delta]^{\frac{1}{\delta} - 1} \times \exp -[(-\ln(u))^\delta + (-\ln(v))^\delta]^{\frac{1}{\delta}}$ (1.84)

L'obtention de $h_G^{-1}(u, v)|_v$ étant complexe, sa valeur est calculée numériquement (voir *Bernard and Czado [2015]*).

1.3.3 Spécification de la R-copule-vigne

1.3.3.1 Algorithme général

Pour spécifier les copules nécessaires pour une construction paire-copule, nous utilisons la structure graphique des R-vignes (voir *Czado* [2019]). L'adaptation d'une spécification de R-vigne copules à un ensemble de données nécessite l'application de certaines étapes, que l'on appelle **Méthode du Maximum spannig tree**. Cette méthode a été détaillée dans *Dißmann et al.* [2013] et peut être résumée dans l'algorithme suivant :

Étape 1 Sélection de la structure R-vigne, c'est-à-dire sélection des paires non conditionnées et conditionnées à utiliser.

Étape 2 Choix d'une famille de copules bivariées pour chaque paire sélectionnée à l'étape précédente.

Étape 3 Estimation du (des) paramètre (s) correspondant (s) pour chaque copule.

La construction du premier arbre de la R-vigne est une étape très importante. L'une des méthodes les plus courantes est celle qui utilise le Tau de Kendall comme poids. D'autres méthodes citées dans *Czado et al.* [2013] telles que l'utilisation des AIC ou des p-values des tests d'ajustement des copules comme poids sont beaucoup moins utilisées.

1.3.3.2 Choix des copules

Après avoir spécifié la structure de la R-vigne adéquate aux données, il faut spécifier à chaque arête du premier arbre les copules adéquates et à chaque arête des arbres suivants les copules conditionnelles adéquates. Le choix des copules bivariées (conditionnelles et inconditionnelles) de chaque arbre se fait individuellement à chaque fois. Les copules sélectionnées sont celles avec le plus petit AIC et leurs paramètres sont estimés par MLE (voir *Aas et al.* [2009]).

La spécification des copules et la dépendance dans les queues est d'une grande importance dans l'estimation conditionnelle des quantiles quand le conditionnement fait intervenir la queue d'une des distributions marginales (voir *Bernard and Czado* [2015]). Ces dernières sont à leur tour des facteurs indispensables dans une distribution à base de copule et doivent tenir compte des queues des échantillons.

Le choix des copules est une étape essentielle dans la spécification des R-vigne copules. Toutefois, il est moins important dans les arbres inférieurs et pour les variables indépendantes que pour les arbres supérieurs et les variables hautement dépendantes. En effet, on remarque qu'à partir

du deuxième arbre, les dépendances deviennent faibles (voir *Bernard and Czado [2015]*). Notons aussi que la diversification du choix des copules d'une structure R-vigne est une stratégie bénéfique, puisque l'utilisation de différentes copules peut réduire le nombre de paramètres, et donc l'erreur d'estimation.

1.3.3.3 Estimation des paramètres

Suivant l'algorithme de spécification des copules-vignes, la méthode d'estimation des paramètres des copules conditionnelles et non conditionnelles constituant les copules-paires de la R-vigne se fait par MLE, de la même façon que pour les copules dans le cas classique. L'application de cette méthode est d'autant plus facile, puisque suivant la définition donnée dans l'équation 1.76, page 33 de la densité d'une R-copule-vigne, sa vraisemblance est un produit de termes, qui se transforme en somme lorsque nous optimisons la *log*-vraisemblance (voir *Czado [2019]*). Nous reprenons ci-après la définition de l'estimation séquentielle des paramètres des R-copules-vignes que nous trouvons dans *Czado [2019]*.

Définition 1.17. Utilisons les notations suivantes :

- Soit θ_e les paramètres de la copule correspondant à l'arête $e = (a_e, b_e; D_e)$ dans une séquence régulière d'arbres de vigne T_i .
- Les paramètres de copules associées aux arêtes de l'arbre T_i sont notés par $\theta(T_i)$ et leurs estimateurs par $\hat{\theta}(T_i)$.
- Nous supposons que tous les paramètres de copules identifiés dans les arbres T_1 à T_{i-1} sont déjà estimés. Nous notons cet ensemble d'estimateurs par $\hat{\theta}(T_1, \dots, i-1)$.

L'estimation séquentielle de θ_e pour l'arête $e = (a_e, b_e; D_e)$ dans l'arbre T_i est basée sur les pseudo-observations

$$u_{k,a_e|D_e, \hat{\theta}(T_1, \dots, i-1)} := C_{a_e|D_e}(u_{k,a_e} | u_{k,D_e}, \hat{\theta}(T_1, \dots, i-1)) \quad (1.85)$$

$$u_{k,b_e|D_e, \hat{\theta}(T_1, \dots, i-1)} := C_{b_e|D_e}(u_{k,b_e} | u_{k,D_e}, \hat{\theta}(T_1, \dots, i-1)), \quad (1.86)$$

pour $k = 1, \dots, n$. En particulier, $\hat{\theta}_e$ est estimé en maximisant

$$\prod_{k=1}^n c_{a_e, b_e; D_e}(u_{k,a_e|D_e, \hat{\theta}(T_1, \dots, i-1)}, u_{k,b_e|D_e, \hat{\theta}(T_1, \dots, i-1)}; \theta_e), \quad (1.87)$$

ou par inversion du τ de Kendall basé sur les pseudo-observations.

1.3.3.4 Comparaison des critères AIC et BIC

Dans le même esprit que pour le choix des copules, l'AIC et le Critère d'Information Bayésien (BIC) sont les critères utilisés pour le choix entre plusieurs structures de R-copules-vignes. Leurs définitions sont données dans *Czado* [2019] comme suit :

Définition 1.18. Pour un échantillon de copule aléatoire u de taille n issu d'une R-copule-vigne avec triplet $(V, B(V), \theta(B(V)))$, l'AIC et le BIC pour la spécification de R-copule-vigne complète sont définis respectivement par

$$AIC_{RV} = -2 \sum_{k=1}^n \ln(l_k(\hat{\theta}(B(V)); u_k)) + 2K, \quad (1.88)$$

$$BIC_{RV} = -2 \sum_{k=1}^n \ln(l_k(\hat{\theta}(B(V)); u_k)) + \ln(n)K. \quad (1.89)$$

Ici K correspond au nombre de paramètres du modèle, c-à-d., la longueur du vecteur $\theta(B(V))$. En plus, $\hat{\theta}(B(V))$ est un estimateur de $\theta(B(V))$.

1.3.3.5 Test d'ajustement de Vuong

Afin de vérifier la qualité de nos ajustements, nous utilisons le test de *Vuong*. Ce test a été introduit par *Vuong* [1989] pour la comparaison de deux structures R-copules-vignes. Bien que la comparaison dans *Brechmann* [2010] de l'efficacité de ce test à celle de l'AIC, du BIC et à celle du test d'ajustement de Cramér-von Mises a révélé la puissance de l'AIC, il reste l'un des outils les plus utilisés pour la comparaison de structures R-vignes, deux à deux. Nous reprenons ci-après la définition de ce test telle que détaillée dans *Dißmann* [2010].

Soit h_0 la vraie densité (inconnue) de la variable aléatoire Y et E_0 son espérance sous le vrai modèle h_0 . On considère les deux modèles $F_\beta = f(Y|\beta)$ et $G_\gamma = g(Y|\gamma)$. Soit β_* et γ_* les vrais paramètres. Alors l'hypothèse nulle du test est

$$H_0 : E_0 \left[\ln \left(\frac{f(Y|\beta_*)}{g(Y|\gamma_*)} \right) \right] = 0 \quad (1.90)$$

H_0 indique que les deux modèles sont également proches de la vraie spécification, c'est-à-dire qu'ils peuvent être considérés comme également bons. Soient $\hat{\beta}_N$ et $\hat{\gamma}_N$ les estimateurs obtenus par

MLE de F_β et G_γ , respectivement. Vuong a montré que sous H_0

$$\hat{V}_N := \frac{\sum_{i=1}^N \ln \left[\frac{f(y_i|\hat{\beta}_N)}{g(y_i|\hat{\gamma}_N)} \right]}{\sqrt{\sum_{i=1}^N \left[\ln \left(\frac{f(y_i|\hat{\beta}_N)}{g(y_i|\hat{\gamma}_N)} \right) \right]^2 - \frac{1}{N} \left[\sum_{i=1}^N \ln \left(\frac{f(y_i|\hat{\beta}_N)}{g(y_i|\hat{\gamma}_N)} \right) \right]^2}} \xrightarrow{D} N(0, 1) \quad (1.91)$$

quand $n \rightarrow \infty$ avec . La p-value correspondante est donnée par $2(1 - \phi(|\hat{V}_N|))$, $\phi(\cdot)$ étant la CDF normale standard. Lorsqu'il compare des modèles avec un nombre différent de paramètres, Vuong suggère d'ajuster la statistique du test du rapport de vraisemblance pour tenir compte du nombre différent de paramètres dans le modèle. En notant $LR_N(\hat{\beta}_N, \hat{\gamma}_N) = \sum_{i=1}^N \ln \left[\frac{f(y_i|\hat{\beta}_N)}{g(y_i|\hat{\gamma}_N)} \right]$, il recommande les deux termes de correction suivants :

$$LR_N^{Akaike}(\hat{\beta}_N, \hat{\gamma}_N) = LR_N(\hat{\beta}_N, \hat{\gamma}_N) - p - q \quad (1.92)$$

$$LR_N^{Schwarz}(\hat{\beta}_N, \hat{\gamma}_N) = LR_N(\hat{\beta}_N, \hat{\gamma}_N) - \frac{p}{2} \ln(N) - \frac{q}{2} \ln(N), \quad (1.93)$$

avec p le nombre de paramètres dans le modèle F_β et q le nombre de paramètres dans G_γ .

1.3.4 Régression quantile

La méthode de régression linéaire des quantiles (voir *Koenker and Bassett Jr [1978]* et *Koenker [2005]*) a été critiquée par *Bernard and Czado [2015]* car elle pose trop de restrictions sur la forme des quantiles de régression. Pour des marges suivant la loi normale, le modèle ne marche pas dès que la structure de dépendance entre la réponse et les covariances dévie d'une copule gaussienne (voir *Kraus and Czado [2017]*). Rappelons la définition de la méthode de régression linéaire des quantiles à travers les deux propositions suivantes de *Koenker [2005]*.

Notons : $[x]^+ = \max\{0, x\}$.

Proposition 1.1. (*Estimation des quantiles*) Soit $\alpha \in (0, 1)$. Soient $(y_i)_{i=1, \dots, N}$ N observations de Y . Alors, une solution à la minimisation convexe suivante

$$\min_{q \in \mathbb{R}} \left\{ \frac{1-\alpha}{N} \sum_{i=1}^N (q - y_i)^+ + \frac{\alpha}{N} \sum_{i=1}^N (y_i - q)^+ \right\} \quad (1.94)$$

existe et est égale au quantile échantillonnal de degré α : $\hat{F}_Y^{-1}(\alpha) = y_{(k)}$ (k ème élément quand l'ordre est croissant) tel que $\frac{k-1}{N} < \alpha \leq \frac{k}{N}$.

Comme démontré dans *Bernard and Czado [2015]*, la solution de cette équation n'est pas unique quand $\alpha = \frac{k}{N}$.

Proposition 1.2. (*Régression quantile conditionnelle*) En supposant que les quantiles conditionnels de $Y|(X_1, X_2, \dots, X_n)$ sont linéaires dans l'ensemble des variables (X_1, X_2, \dots, X_n) , alors le quantile conditionnel résout

$$\min_{\beta \in \mathbb{R}^{n+1}} \left\{ \frac{1-\alpha}{N} \sum_{i=1}^N \left(\beta_0 + \sum_{i=1}^n \beta_i x_{ij} - y_j \right)^+ + \frac{\alpha}{N} \sum_{i=1}^N \left(y_j - \beta_0 - \sum_{i=1}^n \beta_i x_{ij} \right)^+ \right\} \quad (1.95)$$

où pour chaque observation y_j , on observe $(x_{1j}, x_{2j}, \dots, x_{nj})$ pour les n variables utilisées dans le conditionnement.

Comme démontré dans *Chang and Joe* [2019], la régression quantile basée sur les R-copules-vignes a montré sa supériorité sur la régression linéaire en matière d'inférence. L'objectif d'une régression quantile est de prédire le quantile d'une variable de réponse Y étant donné le résultat de certaines variables prédictives X_1, \dots, X_d , ($d \geq 1$), où $Y \sim F_Y$ et $X_j \sim F_j$, $j = 1, \dots, d$. Ainsi, tous les efforts sont concentrés sur la modélisation jointe de Y et X , et plus particulièrement sur la fonction quantile pour α dans $[0, 1]$ (voir *Kraus and Czado* [2017] et *El Adlouni* [2018]).

$$q_\alpha(x_1, \dots, x_d) := F_{Y/X_1, \dots, X_d}^{-1}(\alpha/x_1, \dots, x_d). \quad (1.96)$$

En utilisant des transformées intégrales de probabilité (PIT) (voir *Rosenblatt* [1952] et *Schepsmeier* [2015]) : $V := F_Y(Y)$ et $U_j := F_j(X_j)$, nous avons

$$\begin{aligned} F_{Y/X_1, \dots, X_d}(y/x_1, \dots, x_d) &= P(Y \leq y/X_1 = x_1, \dots, X_d = x_d) \\ &= P(V \leq v/u_1, \dots, u_d) \\ &= C_{V/U_1, \dots, U_d}(v/u_1, \dots, u_d), \end{aligned} \quad (1.97)$$

l'inversion donne

$$F_{Y/X_1, \dots, X_d}^{-1}(\alpha/x_1, \dots, x_d) = F_Y^{-1}(C_{V/U_1, \dots, U_d}^{-1}(\alpha/u_1, \dots, u_d)). \quad (1.98)$$

Une estimation du quantile conditionnel peut donc être obtenue en estimant les marges F_Y, F_j ($j = 1, \dots, d$) ainsi que la copule $C_{V/U_1, \dots, U_d}$ (voir *Kraus and Czado* [2017]) comme suit :

$$\hat{q}_\alpha(x_1, \dots, x_d) := \hat{F}_Y^{-1}(\hat{C}_{V/U_1, \dots, U_d}^{-1}(\alpha/\hat{u}_1, \dots, \hat{u}_d)). \quad (1.99)$$

CHAPITRE 2

DÉTECTION DU CHANGEMENT CLIMATIQUE ET PRÉVISION DE L'ENVASEMENT PAR DES COPULES

Introduction

Le changement climatique est un défi mondial nécessitant les efforts joints de tous les pays. Il est défini universellement par le Groupe d'experts intergouvernemental sur les changements climatiques (IPCC) comme étant « un changement de l'état du climat qui peut être identifié (par exemple à l'aide de tests statistiques) par des changements dans la moyenne et/ou la variabilité de ses propriétés, et qui persiste pendant une période prolongée, généralement des décennies ou plus. Il fait référence à tout changement du climat au fil du temps, que ce soit en raison de la variabilité naturelle ou en raison de l'activité humaine».

L'état hydrologique étant l'une des composantes principales du climat, acquiert aussi une attention particulière. Le stress hydrique est en effet l'un des sujets les plus concernés par les études que ce soit du point de vue économique ou purement environnemental. De plus en plus de recommandations s'orientent vers la construction de nouvelles réserves d'eau (barrages) pour palier au problème de la sécheresse, des précipitations annuelles non régulières mais aussi pour répondre à la demande énorme que subit l'eau dans tous les secteurs (industries diverses, agriculture basée sur l'irrigation, consommation en eau potable, production d'énergie). Mais avant de débattre la nécessité ou non de construire un barrage dans telle ou telle région, il faut s'assurer de l'état des constructions déjà existantes, notamment s'assurer de l'utilisation optimale de leurs capacités de stockage et estimer la quantité exacte d'eau qu'ils contiennent.

De nombreuses recherches ont été menées pour analyser Les effets du changement climatique sur l'hydrologie. Pour ne citer que quelques exemples de recherches intéressantes sur le sujet, les travaux de *Driouech et al.* [2010] et *Tekken and Kropp* [2012] étudient l'impact du changement climatique sur les variables pluviométriques et de ruissellement et sur la pénurie d'eau au Maroc, celui de *Aguilera et al.* [2015] traite de cet impact sur la dynamique des nutriments des rivières et le document de *Yagbasan* [2016] discute de cet effet sur la recharge des eaux souterraines dans la région méditerranéenne. Les travaux de *Hosseinzadeh et al.* [2015], *Les Landes et al.* [2014] et *Yu et al.* [2016] étudient l'impact du changement climatique sur la sécheresse, l'extension des zones humides et les ressources en eau, respectivement. Dans *Driouech et al.* [2010], on utilise des méthodes de réduction d'échelle spatiale (downscaling) statistique et dynamique et des données de scénarios climatiques à haute résolution pour évaluer l'impact du changement climatique sur les ressources en eau du bassin de la Moulouya pour l'horizon 2021-2050.

Les ressources en eau sont particulièrement affectées par le changement climatique récent en termes de changements inter-annuels des régimes pluviométriques, de diminution des quantités de

précipitations, d'augmentation des températures et de cycles de sécheresse persistants (voir *Bates et al.* [2008]). Ainsi, la combinaison d'une demande croissante en eau et d'une disponibilité en eau de plus en plus limitée devrait entraîner une grave pénurie d'eau à l'avenir, à moins que des approches et des contre-mesures innovantes ne soient prises (voir *Jackson et al.* [2001]).

À l'échelle marocaine, un rapport de la Banque mondiale (voir *Dahan and Grijsen* [2017]) indique que le Maroc sera confronté à un stress hydrique extrêmement élevé en 2040 (la surexploitation des eaux souterraines atteint en moyenne près d'un milliard de m^3 par an). En plus de cette carence, la croissance démographique et les effets du changement climatique font que les parts individuelles de l'eau retombent à environ $700 m^3$ par an. Selon le même rapport, en 1960, la part individuelle de l'eau était de $3500 m^3$. Dans ces conditions difficiles, la gestion des ressources en eau et des réserves est devenue une urgence. L'entretien des barrages est l'une des pratiques proposées pour préserver la qualité de l'eau et maximiser son stockage.

Un facteur majeur affectant l'évaluation de la quantité réelle d'eau dans un barrage est l'envasement. C'est le résultat d'un processus complexe que les sédiments suivent pour se déposer dans un réservoir de barrage. L'envasement varie en fonction de plusieurs variables, notamment les précipitations, la topographie, le couvert végétal, les cultures, les parcours, les forêts et le bassin versant (voir *Morris and Fan* [1998]). D'autres facteurs viennent s'ajouter à ces variables, notamment le changement de température qui a des effets prépondérants sur l'altération mécanique des roches et le vent, qui est un autre agent important d'érosion (voir *Vanoni* [2006]). Mais l'influence de ces deux variables climatiques sur l'érosion est difficile à évaluer à cause de leur variation (voir *Benaïcha* [2011]). Cette multi-dépendance rend la modélisation de l'envasement une tâche difficile puisque les données historiques de ces variables sont souvent indisponibles. En général, le suivi du niveau d'envasement est effectué par un relevé bathymétrique de la zone autour de l'ouvrage de drainage de fond (voir les différentes techniques utilisées pour réaliser un relevé bathymétrique dans la Figure 2.1). Cette opération ne peut être menée régulièrement par les autorités en raison de son coût très élevé, ce qui entraîne des conclusions trompeuses et des décisions erronées en matière de gestion des ressources en eau. Par conséquent, l'élaboration d'un modèle mathématique d'envasement s'avère plus que nécessaire.

La première description analytique du processus d'envasement a été faite par *Orth* [1934] (cité dans *Saidi* [1991]), où il est indiqué que la capacité résiduelle diminue de façon exponentielle avec le temps : Après t années, la capacité résiduelle C_t est donnée par $C_t = C_{Rn} \left(1 - \frac{Q_{st}}{C_{Rn}}\right)^t$ où C_{Rn} est la capacité normale et Q_{st} l'apport annuel moyen en solides. D'autres méthodes géométriques bien connues d'estimation de l'envasement telles que la méthode de Kolmogorov et la méthode de la hauteur moyenne sont basées sur la mesure directe du taux d'envasement en visitant la zone

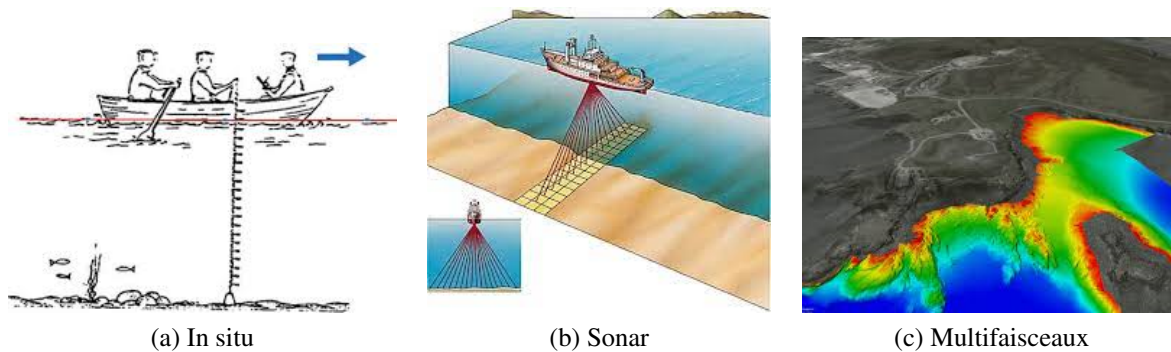


FIGURE 2.1 – Quelques techniques de réalisation de relevés bathymétriques.

d'étude sans utiliser la modélisation et les formules empiriques (voir *Alahiane et al.* [2014]). Une autre méthode proposée dans *Remini and Hallouche* [2005] utilise un modèle de régression linéaire pour établir une relation entre l'envasement et l'âge des barrages suivant la force de sédimentation puis utilise ces résultats pour la prévision de l'envasement dans les barrages du Maghreb.

Dans cette partie de la thèse, nous présentons et discutons les résultats de l'application de notre méthodologie aux données du barrage Mohammed V situé au bassin de la Moulouya. Ce bassin est connu pour son rôle de pilote dans le développement de l'économie et de l'agriculture de la région orientale du Maroc. Il fait face à un stress croissant sur ses ressources en eau pour répondre aux besoins des grandes activités agricoles irriguées (agrumes, céréales et oliviers). Nous cherchons à détecter les effets du changement climatique sur les variables hydrologiques et à prévoir l'envasement à partir de variables climatiques mesurables à l'aide des copules. Nous montrons également comment la copule la mieux adaptée à nos variables est retenue pour prévoir l'envasement de la prochaine décennie suivant différents scénarios de changement de température. Notre approche utilise des données climatiques qui ont le grand avantage d'être largement, voire publiquement, disponibles et régulièrement mises à jour. De ce fait, il est possible de surmonter les limites des autres modèles analytiques, et de pouvoir fournir des estimations en temps réel de l'envasement d'un barrage.

Ce chapitre se présente comme suit : Après une présentation de la région d'étude et des variables en usage, nous détaillons les arguments appuyant la détection de l'effet du changement climatique sur les variables hydrologiques dans la section 2.3 et nous présentons les étapes de l'obtention de notre modèle de prévision de l'envasement dans la sous-section 2.3.6. Nous terminons cette partie par une conclusion où nous résumons les principaux résultats et discutons les principales difficultés rencontrées auxquelles nous proposons des alternatives.

2.1 Région d'étude

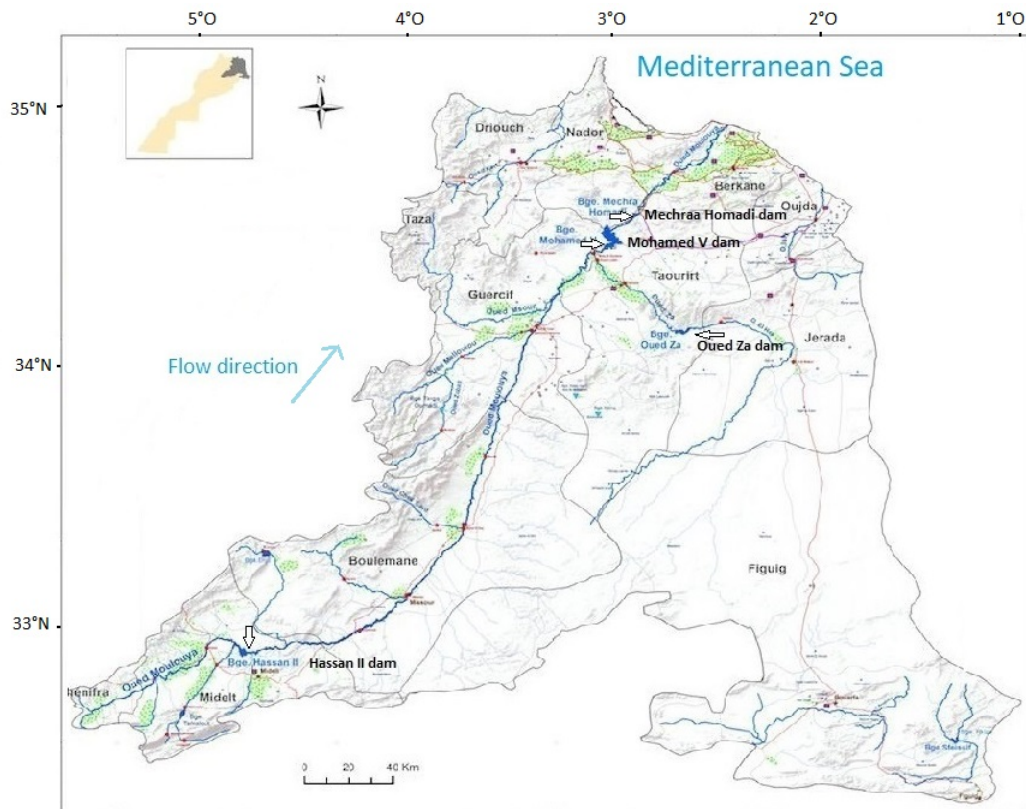
Le bassin de la Moulouya (Figure 2.2) ($35^{\circ} 07' 26''$ N, $2^{\circ} 20' 30''$ O) est le plus grand bassin du Maroc et des fleuves non sahariens d'Afrique du Nord avec une superficie de $54\,500\text{ km}^2$. La rivière de la Moulouya, longue d'environ 600 km, prend sa source à Alemsid à une altitude de 1170 m à la jonction du massif du moyen Atlas et du haut Atlas et se jette à Saidia (Ras el ma) en Méditerranée (voir *Driouech et al.* [2010]). L'apport moyen annuel à l'embouchure de la Moulouya est de l'ordre de 920 millions de m^3/an (période 1970-1998). L'émissaire (le canal d'évacuation) et sa zone marécageuse représentent un site d'intérêt biologique et écologique s'étendant sur environ 3000 hectares (voir *Driouech et al.* [2010]). Le bassin joue un rôle important dans le développement de l'économie et de l'agriculture de la région orientale du Maroc.

La rivière Moulouya est utilisée pour l'irrigation à travers cinq barrages (Mohammed V, Machraa Hammadi, Hassan II, Enjil et Arabat). Son bassin versant peut être divisé en trois bassins : le bassin de la Haute Moulouya, au sud, se prolonge par le bassin de la Moyenne Moulouya. Celle-ci continue vers le bassin inférieur de la Moulouya (voir *Diani et al.* [2017]).

Le bassin de la Haute Moulouya est un bassin intramontagnard de forme triangulaire. Il est bordé à l'ouest par le Moyen Atlas et au sud par le Haut Atlas. Il forme un bassin recevant des matériaux sédimentaires issus du démantèlement du Haut Atlas et du Moyen Atlas. Sa physiographie est dominée par des plaines (85%) limitées par des montagnes (15%). En plaine, les altitudes sont inférieures à 2000 m alors qu'en montagne, elles dépassent 2000 m (voir *Diani et al.* [2017]).

La zone hydrographique de la Moulouya se trouve sous une double influence climatique : Atlantique et Méditerranéenne (voir *Driouech et al.* [2010]). Elle se caractérise par un climat sec et aride avec des tendances vers un climat méditerranéen au Nord-Est (en Basse Moulouya et côtes méditerranéennes), continental en Moyenne Moulouya et montagneux en Haute Moulouya. Les précipitations sont généralement comprises entre 100 et 400 mm / an sur la majeure partie du bassin et atteignent 700 mm / an dans les monts Horsts, Béni Bou Yahy-Béni Snassène et dans la zone Nord-Est (bassins côtiers et Rif oriental). Les températures moyennes du bassin varient généralement entre 5°C et 18°C en hiver et entre 18°C et 31°C en été (voir *Driouech et al.* [2010]).

Des informations complémentaires sur le bassin sont disponibles sur le site de l'Agence du bassin hydrologique de la Moulouya et dans *Snoussi* [2004].



(a)

Source : Agence du bassin hydrologique de la Moulouya (modifié).



(b)

Source : Agence du bassin hydrologique de la Moulouya.

FIGURE 2.2 – Barrage Mohammed V dans le bassin de la Moulouya.

2.2 Données

Les données disponibles sont des observations de séries chronologiques de trois types :

TABLE 2.1 – *Bathymétrie annuelle (Mm^3) au barrage Mohammed V.*

Mesures bathymétriques annuelles (Mm^3 /an) du barrage Mohammed-V.		Année	Volume actif	Envasement cumulé
Type 1	Nous disposons de 11 observations dispersées entre 1975 et 2011 présentées dans le Tableau 2.1, les valeurs annuelles restantes sont simulées à l'aide du package Multiple Imputation by Chained Equations (MICE) du logiciel R (voir annexe A.3, page 88 pour plus de détails concernant cette méthode).	1967	725,75	0,00
		1975	595,68	130,07
		1985	489,24	236,51
		1990	467,72	258,03
		1994	410,92	314,82
		1997	403,21	322,54
		2000	383,30	342,45
		2001	362,55	363,19
		2002	331,58	394,17
		2003	326,69	399,06
		2008	279,39	446,36
	2011	239,52	486,22	

Source : Ministère de l'Équipement du Transport, de la Logistique et de l'Eau.

Type 2 Apports en eau (Mm^3 /saison de l'automne) au barrage Mohammed-V. Nous utilisons les valeurs cumulées des observations d'octobre, novembre et décembre pour représenter la saison d'automne.

Type 3 Données climatiques :

1. Températures en $^{\circ}C$, mesurées à 2 mètres au-dessus de la surface de la terre. Nous considérons la moyenne estivale de : la température moyenne quotidienne, la température maximale quotidienne et la différence entre les températures maximales et minimales quotidiennes.
2. Précipitations automnales moyennes (mm / jour), et
3. Vitesse moyenne du vent en automne mesurée à 2 mètres au-dessus de la surface de la terre (m / s). Notons, au passage, que pour les études liées au sol (et automatiquement l'érosion et l'envasement), la hauteur recommandée pour la mesure de la vitesse du vent est de 2 m au-dessus de la surface (voir *Zobeck and Van Pelt [2005]*).

Les données climatiques disponibles sur le site de la NASA ne le sont qu'à partir de 1981. Les valeurs manquantes (de 1968 à 1980) sont simulées à l'aide du package R susmentionné.

Notons ici que même si la vitesse maximale du vent semble plus adaptée à notre étude, une étude récente de *Colosimo et al. [2020]* montre que même des vents modérés, lorsqu'ils sont opposés aux

vents dominants, peuvent inverser le courant et générer un transport résiduel important. De plus, selon la même étude, lorsque le débit est faible, même des vents faibles peuvent influencer le débit net. D'autre part, un maximum ne peut être observé qu'une seule fois dans la saison, tandis que des valeurs voisines peuvent être observées plusieurs fois et ont donc un impact plus important. La moyenne présente l'avantage d'être sensible à ces valeurs extrêmes.

Pour tout ce qui suit, notons par $(W_t)_{t \geq 0}$, $(P_t)_{t \geq 0}$, $(\bar{T}_t)_{t \geq 0}$, $(T_{max_t})_{t \geq 0}$, $(T_{dt})_{t \geq 0}$, $(WS_t)_{t \geq 0}$ et $(S_t)_{t \geq 0}$ les séries chronologiques des variables apport en eau, précipitations, température moyenne, température maximale, différence de température diurne, vitesse du vent et envasement, respectivement.

2.2.1 Exploration graphique

Nous représentons sur la Figure 2.3 les séries de données des sept variables étudiées après avoir remplacé les données manquantes par leurs valeurs estimées.

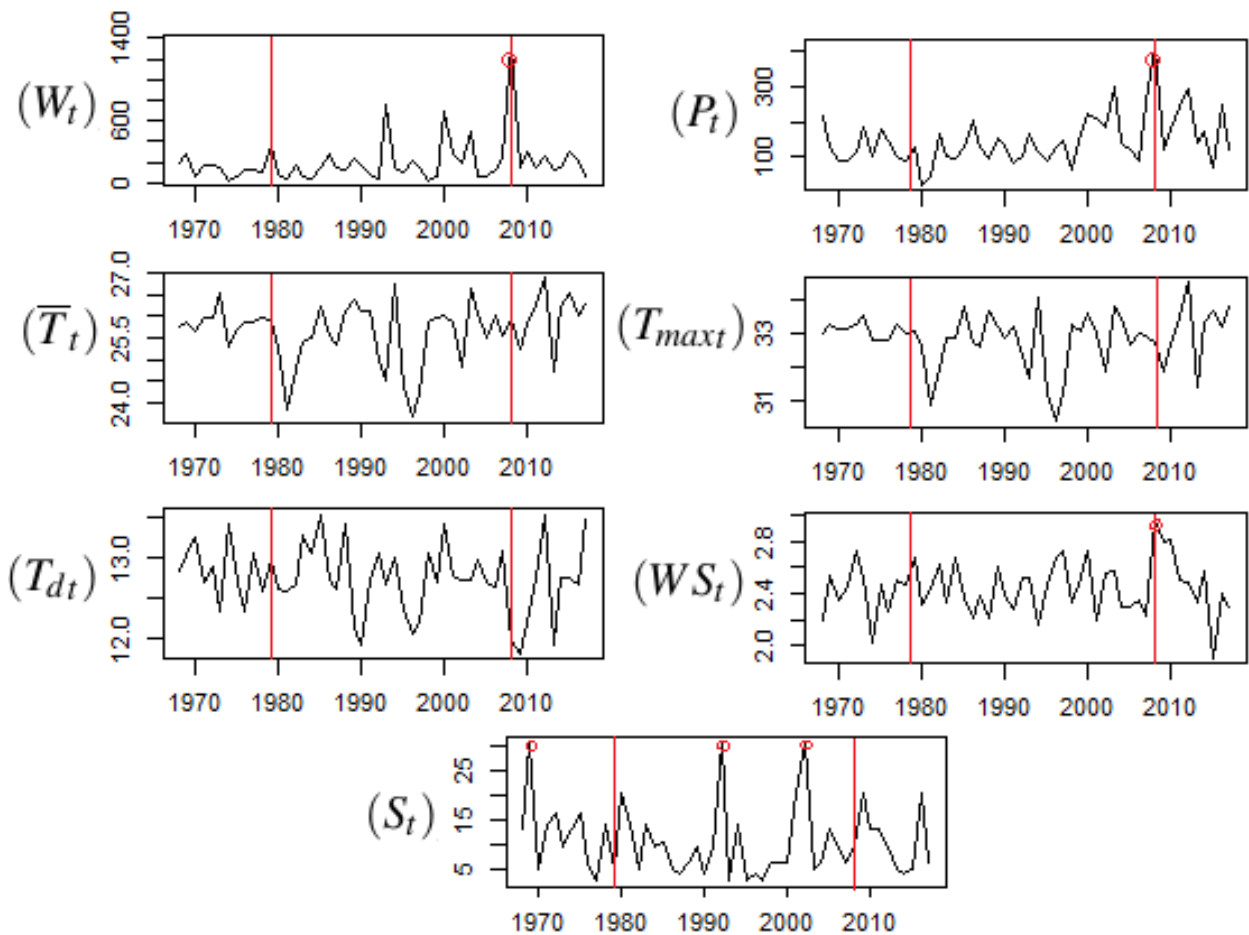


FIGURE 2.3 – Séries chronologiques des variables étudiées.

Pour la série W_t , nous remarquons à partir de l'année 1993 une tendance à la hausse et une

variabilité plus prononcée avec un pic dépassant 1300 Mm^3 en 2008. La série P_t présente aussi une certaine tendance à partir de 2000 avec un pic en 2008 aussi. Pour les séries de température, on remarque une grande variabilité dans la variance avec une claire analogie de comportement des séries \bar{T}_t et $T_{max,t}$. Pour la série WS_t , il n'y a pas de tendance générale et les variations apparaissent comme d'amplitude constante, quoiqu'on remarque la présence d'un pic de 3 m/s en 2008. Finalement, pour la série de l'envasement S_t , nous remarquons une variabilité modérée sans tendance particulière avec des pics en 1969 (2ème année après le démarrage du barrage), en 1992 et en 2002 dépassant les 25 Mm^3 . Les pics détectés ont été encerclés en rouge pour être mieux visibles (voir annexe A.4 Tableau A2, page 90 pour les statistiques descriptives de ces données).

2.2.2 Choix des périodes

Comme indiqué dans les études de *Singla* [2009] et *Driouech et al.* [2010] concernant la relation entre le changement climatique et le régime hydrologique au Maroc portant sur des données jusqu'en 2008, le comportement hydrologique des deux stations Dar-El Caïd et Zaida représentant la région de Moulouya a connu un changement brutal en 1979. Ce changement comprend les régimes de précipitations et de ruissellement. Puisque la région n'a pas connu d'interventions significatives cette année-là, les deux études indiquent que le changement climatique est le principal responsable de ce changement hydrologique.

Partant de cette hypothèse, nous avons scindé les séries des données observées en trois périodes distinctes : De 1968 à 1978 (que nous appellerons période 1), de 1979 à 2008 (période 2) et de 2009 à 2017 (période 3). Les limites entre ces trois périodes sont représentées par des traits rouges verticaux sur la Figure 2.3. Pour les deux premières périodes, nous comparerons nos résultats avec ceux de *Singla* [2009] et *Driouech et al.* [2010]. A partir de 2009, nous appliquerons notre méthode sans référence aux études précédentes. Pour chacune des trois périodes, nous construisons les copules reliant l'apport en eau et les variables climatiques, celles de l'envasement et les variables climatiques puis celles de l'envasement et de l'apport en eau. La comparaison des copules obtenues dans les deux premières périodes est ensuite utilisée pour prouver l'éventuel impact du changement climatique sur le régime hydrologique de la région. Ensuite, nous regroupons les données de la période 2 et de la période 3 (que nous appellerons période 23) pour examiner si une nouvelle occurrence du changement de comportement hydrologique a eu lieu. Enfin, en utilisant des techniques d'ajustement, toutes les copules obtenues sont examinées pour choisir celle qui servira pour construire un modèle de l'envasement de la période 3.

2.2.3 Analyse des corrélations

Pour explorer les relations de dépendance entre nos variables, nous calculons les coefficients de corrélation de Kendall et Spearman. Le Tableau 2.2 montre ces coefficients à différentes périodes. Les résultats montrent que pendant la période 1, l'apport en eau est corrélé aux précipitations, aux températures moyennes et maximales et à la vitesse du vent. Pour l'envasement, les seules corrélations remarquables sont celles avec la vitesse du vent et l'apport en eau.

TABLE 2.2 – *Mesures de corrélation entre les variables.*

X_i	X_j	Période 1		Période 2		Période 3		Période 23	
		Kendall	Spearman	Kendall	Spearman	Kendall	Spearman	Kendall	Spearman
W	P	0.24	0.32	0.58	0.76	0.39	0.37	0.51	0.68
W	\bar{T}	0.31	0.48	0.09	0.12	0.28	0.37	0.13	0.18
W	T_{max}	0.45	0.61	0.02	0.02	0.17	0.18	0.06	0.09
W	T_d	-0.13	-0.18	-0.14	-0.20	0.00	0.05	-0.10	-0.15
W	WS	0.31	0.41	0.20	0.30	0.11	0.22	0.19	0.28
S	P	0.07	0.07	-0.17	-0.24	0.26	0.40	-0.06	-0.08
S	\bar{T}	0.04	0.13	0.08	0.13	-0.15	-0.21	0.04	0.08
S	T_{max}	0.04	0.07	0.10	0.17	-0.15	-0.23	0.04	0.09
S	T_d	0.00	-0.10	0.23	0.32	-0.20	-0.29	0.13	0.16
S	WS	0.30	0.41	-0.21	-0.32	0.32	0.41	-0.07	-0.10
S	W	0.30	0.34	-0.22	-0.35	0.09	0.18	-0.16	-0.24

La corrélation entre l'envasement et la différence de température diurne, imperceptible en période 1, s'est renforcée au cours de la période 2, alors que la corrélation entre les variables de l'apport en eau et de température n'est plus perceptible. L'augmentation de la taille des données peut être derrière ce changement. Le Tableau 2.2 montre également que l'apport en eau et les précipitations restent fortement corrélés, tandis que l'envasement devient négativement corrélé avec l'apport en eau et la vitesse du vent. On note aussi que certaines petites dépendances dans la période 2 deviennent perceptibles dans la période 3. C'est le cas de la corrélation entre l'apport en eau et les variables de température. Les résultats de la période 23 sont presque similaires à ceux de la période 2 en raison de la prédominance des données de cette dernière.

Nous explorons aussi les corrélations se produisant entre deux séries Y_t et $X_{t\pm k}$ où $\pm k$ est un décalage. Les résultats sont représentés dans le Tableau A1 (voir Annexe A.4.2).

2.2.4 Modélisation ARMA

Avant d'entamer la modélisation ARMA, nous avons procédé à la transformation des séries étudiées. Suivant l'approche de *Chan et al.* [1977], nous avons utilisé le *log* pour stabiliser les variations au cours du temps et la différenciation simple afin de supprimer le mouvement tendanciel (voir

Annexe A.4, Tableau A3, page 92 pour les résultats des tests de stationnarité Dickey-Fuller augmenté (méthode dans *Said and Dickey* [1984]) et KPSS (méthode dans *Kwiatkowski et al.* [1992]) avant et après les transformations). Pour garder une uniformité dans l'interprétation des résultats, toutes les séries ont subi les mêmes transformations. Tous nos processus étant stationnaires, nous pouvons utiliser une modélisation par un processus ARMA (voir *Lagnoux* [2015b]). Nous présentons dans le Tableau 2.3 les modèles ARMA ajustés et les résultats des tests de normalité et d'indépendance de leurs résidus.

TABLE 2.3 – Modèles ARMA retenus et résultats des tests de normalité et d'indépendance des résidus.

Variable	Période 1			Période 2				Période 3		Période 23			
	ϕ_1	θ_1	σ_ε^2	ϕ_1	ϕ_2	θ_1	σ_ε^2	ϕ_1	σ_ε^2	ϕ_1	ϕ_2	θ_1	σ_ε^2
<i>W</i>	0	-0.828	0.544	0.107	-0.387	-0.955	0.894	0	0.466	0	0	-0.983	0.781
<i>P</i>	0	0	0.154	-0.332	-0.733	0	0.270	-0.653	0.312	0	0	-0.819	0.306
\bar{T}	-0.533	0	0.0002	-0.359	-0.403	0	0	0.001	0.002	0	0	-0.870	0.001
T_{max}	0	0	$9.92 \cdot 10^{-5}$	-0.360	-0.421	0	0	0.001	0.002	0	0	-0.928	0.001
T_d	-0.601	0	0.001	-0.478	-0.318	0	0	0.001	0.004	-0.463	-0.306	0	0.002
<i>WS</i>	0	0	0.016	-0.813	-0.595	0	0.008	-0.609	0.014	-0.713	-0.374	0	0.010
<i>S</i>	0	-0.889	0.562	-0.754	-0.381	0	0.613	0	0.521	-0.682	-0.339	0	0.594

Variable	Période 1			Période 2			Période 3			Période 23		
	K-S	A-D	B-P	K-S	A-D	B-P	K-S	A-D	B-P	K-S	A-D	B-P
<i>W</i>	0.1957	0.3593	0.8023	0.9580	0.9485	0.8661	0.9805	0.9620	0.9876	0.9017	0.9000	0.4792
<i>P</i>	0.9711	0.9282	0.9637	0.4396	0.6042	0.4352	0.9256	0.9366	0.8773	0.8765	0.9267	0.3760
\bar{T}	0.3889	0.6491	0.9383	0.9958	1.0000	0.8582	0.4518	0.6318	0.9022	0.3029	0.3104	0.9285
T_{max}	0.7268	0.8400	0.5720	0.5215	0.8859	0.8901	0.5077	0.5564	0.9023	0.4018	0.5179	0.9544
T_d	0.9919	0.9995	0.7505	0.9953	0.9973	0.5331	0.7188	0.5439	0.8776	0.9975	0.9980	0.5073
<i>WS</i>	0.9392	0.9759	0.7051	0.6254	0.4569	0.9333	0.9742	0.9603	0.9286	0.9158	0.9659	0.8923
<i>S</i>	0.5240	0.7481	0.8660	0.9541	0.9294	0.5390	0.6838	0.8303	0.9739	0.9878	0.9633	0.4179

2.3 Résultats

2.3.1 Analyse graphique des résidus

2.3.1.1 Nuages de points

Nous commençons cette partie de l'analyse par une exploration graphique de la structure de dépendance entre les résidus. Nous représentons sur la Figure 2.4 des exemples des nuages de points obtenus pour les résidus du modèle de l'envasement et celui des autres variables. Tous les nuages de points représentés semblent dispersés et n'indiquent pas de direction forte de la relation entre les variables. Nous pouvons toutefois distinguer une relation négative modérée entre les résidus de l'envasement et l'apport en eau (sous-figure 2.4a), légèrement négative entre le résidu de l'envasement et chacun des résidus des précipitations et du vent (sous-figures 2.4c et 2.4d) ainsi qu'une relation positive très faible entre le résidu de l'envasement et celui de la température maximale (sous-figure 2.4b).

2.3.1.2 K-plots

Nous présentons sur la Figure 2.5, différents exemples des K-plots obtenus pour les résidus normalisés des variables aléatoires étudiées. Sur la sous-figure 2.5a représentant le K-plot de l'apport en eau et les précipitations pendant la période 23, nous pouvons voir que les points (W_i, H_i) sont assez proches de la courbe de dépendance positive complète. Ceci suggère une dépendance positive assez importante, confirmée par les valeurs du τ de Kendall (0,51) et celle du ρ de Spearman (0,68). Une copule Archimédienne représentant les fortes dépendances positives telle que la copule de Gumbel pourrait être choisie dans ce cas. Sur les sous-figures 2.5b et 2.5c représentant les K-plots de l'apport en eau avec la vitesse du vent puis avec la température moyenne pendant la période 23, la disposition des points montre aussi un écart net par rapport à la diagonale, quoique moins prononcé que celui de la sous-figure 2.5a, laissant suggérer une dépendance positive modérée à faible.

D'autre part, la sous-figure 2.5d représentant le K-plot de l'envasement et la température maximale pendant la période 2 montre l'alignement presque parfait des points (H_i, W_i) sur la ligne d'indépendance, les valeurs du τ de Kendall et du ρ de Spearman sont effectivement très faibles (0,10 et 0,17) pour ces deux variables. Un test d'indépendance (voir le paragraphe 1.2.4) devrait être conduit dans ce cas. Les sous-figures 2.5e et 2.5f, représentant les K-plots de l'envasement avec la vitesse du vent puis avec l'apport en eau, montrent l'écart vers le bas des points représentés par rapport à la diagonale, ce qui suggère la présence d'une association légèrement négative dans les

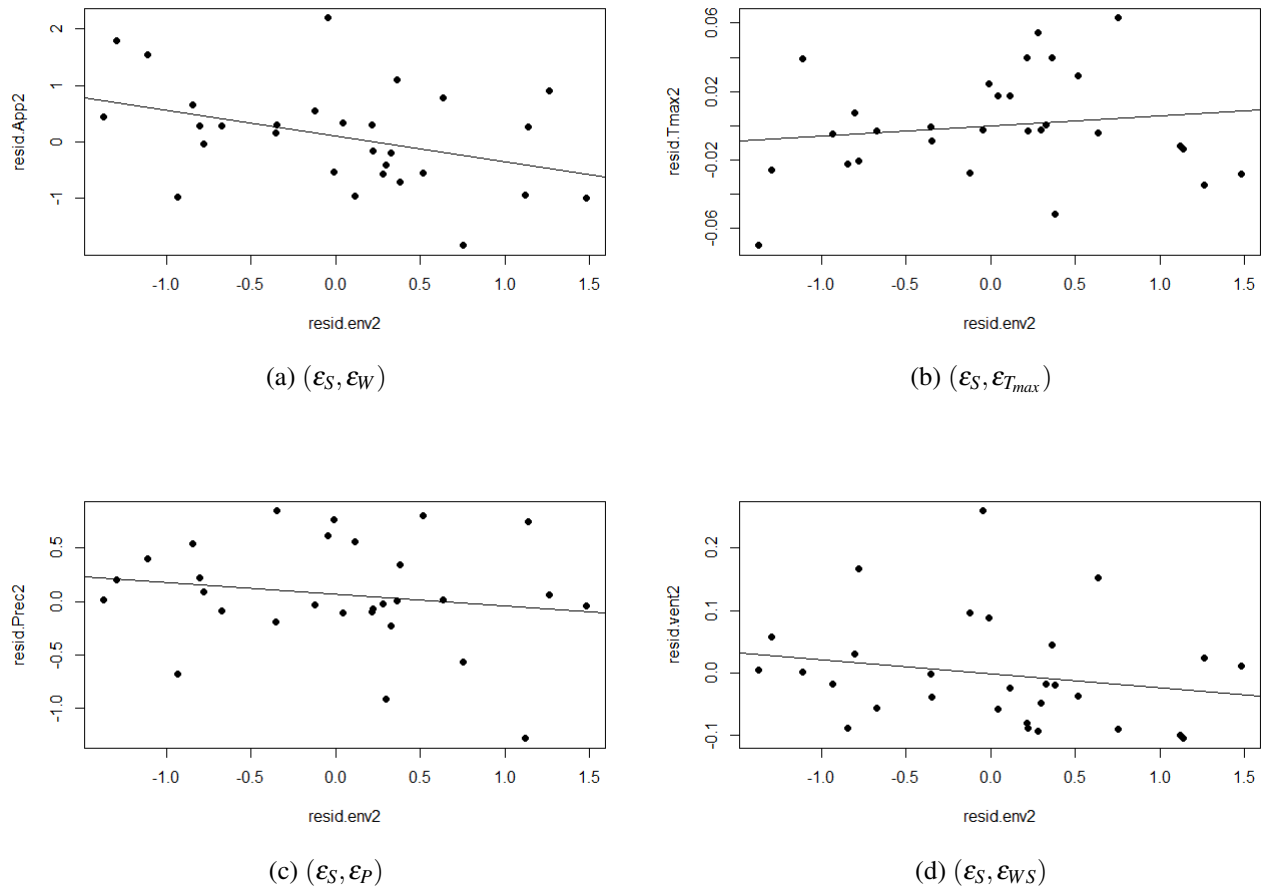


FIGURE 2.4 – Nuages de points entre les couples de résidus de la période 2.

données ce qui est en accord avec les valeurs $\tau=-0,21$, $\rho=-0,32$ pour la première paire et $\tau=-0,22$, $\rho=-0,35$ pour la seconde. Pour ces deux derniers cas, les copules choisies devraient figurer parmi les copules tournées de 90 ou 270 degrés (voir le paragraphe 1.1.4).

2.3.2 Estimation des paramètres

Dans cette étape, toutes les copules candidates (Clayton, Gumbel, Frank, Joe et Tawn) et leurs rotations seront ajustées aux paires transformées (U, V) des variables aléatoires (X, Y) obtenues suivant la description donnée dans le paragraphe 1.2.1, page 24. Nous présentons dans le Tableau A4 (annexe A.5) page 93 les résultats de l'estimation des paramètres des copules par MLE pour les quatre périodes d'étude.

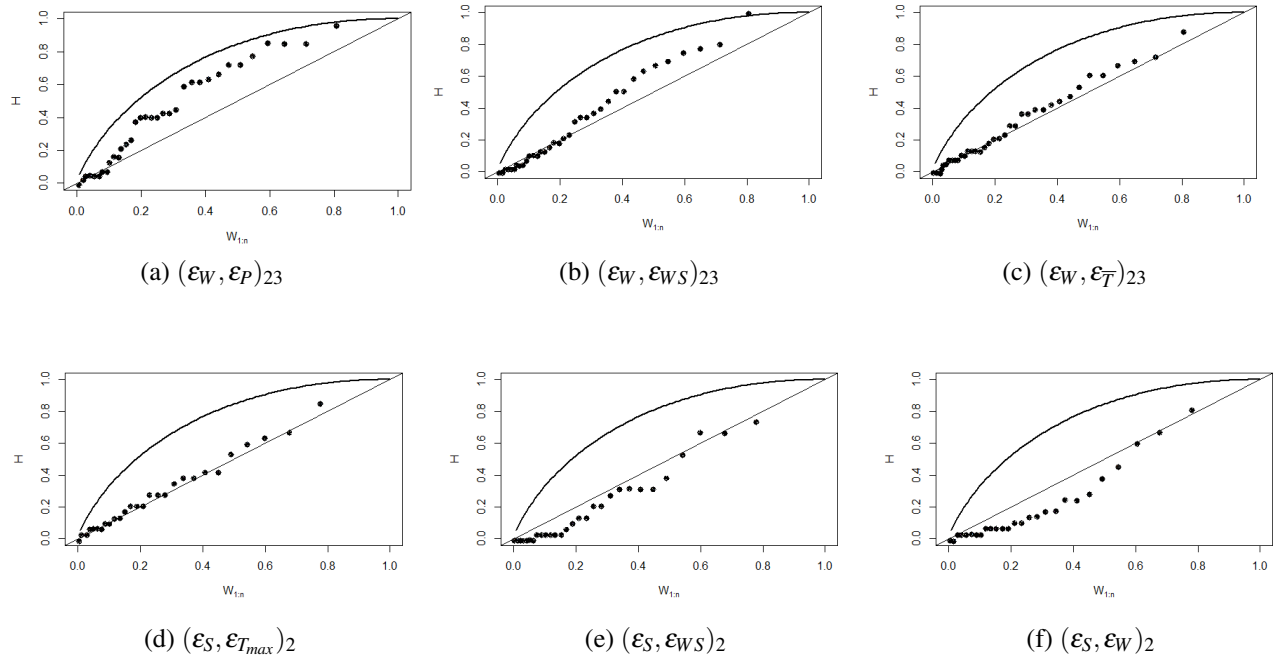


FIGURE 2.5 – Exemples de K-plots obtenus.

2.3.3 Comparaison des critères AIC

Pour chacune des copules résultant de l'étape précédente, nous calculons l'AIC donné par l'équation 1.63, page 29. Nous retenons pour chaque paire de variables la copule avec la plus petite valeur du AIC. Les résultats des différentes périodes sont représentés dans le Tableau A5 (annexe A.5), page 94 où les copules retenues sont données en gras.

2.3.4 Résultats du test d'ajustement

Dans le Tableau 2.4, nous présentons les copules retenues suivant le critère AIC et les p-values correspondantes du test de qualité d'ajustement de Cramér-von Mises.

2.3.5 Détection du changement climatique par les copules

En comparant les résultats du Tableau 2.4 pour les périodes 1 et 2, nous remarquons une transition majeure dans la copule liant l'envasement et la vitesse du vent qui est passée de la copule de Clayton à la copule de Tawn. Ce passage d'une copule de famille Archimédienne à une autre à valeurs extrêmes est illustré par le changement complet du contour de densité de copule représentant le comportement joint de l'envasement et de la vitesse du vent sur les sous-figures 2.6a et 2.6b. Ce

TABLE 2.4 – Résultat du test d'ajustement de Cramér-von Mises des copules retenues.

U_i	U_j	Period 1		Period 2		Period 3		Period 23	
		Copula	p	Copula	p	Copula	p	Copula	p
U_W	U_P	r90 T2(-7.41,0.36)	0.17	Sur T1(4.6,0.3)	0.87	r270 T2(-20,0.24)	0.75	Gumbel(1.98)	0.40
U_W	$U_{\overline{T}}$	Sur Joe(3.38)	0.99	r90 Joe(-1.27)	0.95	Sur T2(20,0.3)	0.66	Sur Gumbel(1.13)	0.55
U_W	$U_{T_{max}}$	Clayton(2.99)	0.48	r90 Joe(-1.39)	0.54	Sur T1(20,0.28)	0.54	Sur Joe(1.11)	0.68
U_W	U_{T_d}	r90 Joe(-2.23)	0.13	r90 T1(-3.28,0.53)	0.19	r90 T1(-20,0.14)	0.72	r90 Gumbel(-1.29)	0.58
U_W	U_{WS}	Sur Joe(2.76)	0.81	Joe(1.89)	0.95	r90Joe(-2,22)	0.13	Joe(1.52)	0.56
U_S	U_P	T1(3.32,0.4)	0.89	r270 T2(-2.28,0.22)	0.46	T1 (5.02,0.49)	0.66	r270 Gumbel(-1.17)	0.62
U_S	$U_{\overline{T}}$	r270 T1(-5.44,0.33)	0.66	Sur T2(2.23,0.3)	0.97	T1(20,0.2)	0.95	Sur Joe(1.32)	0.75
U_S	$U_{T_{max}}$	Independence(0)	0.65	Sur T2(2.31,0.29)	0.66	Independence(0)	0.80	Sur Joe(1.29)	0.20
U_S	U_{T_d}	Sur T1(6.27,0.31)	0.64	T1(20,0.2)	0.60	r270 T1(-4.64,0.41)	0.28	Sur Joe(1.32)	0.66
U_S	U_{WS}	Sur Clayton(1.43)	0.52	r270 T2(-3.47,0.34)	0.44	r270 T2(-9.08,0.49)	0.42	r270 Gumbel(-1.03)	0.70
U_S	U_W	Independence(0)	0.65	r270 Gumbel(-1.48)	0.56	Sur Joe(2.27)	0.89	r270 Gumbel(-1.3)	0.40

changement peut s'expliquer par le rôle croissant que joue la vitesse du vent dans le charriage des sédiments conduisant à une augmentation de l'envasement dans le barrage.

Un autre changement majeur est celui de la copule liant l'apport en eau et la température maximale que l'on peut observer sur les sous-figures 2.6d et 2.6e par un passage de la copule de Clayton à la copule r90 Joe. Même si la transition reste dans la même famille Archimédienne, elle révèle une dépendance plus forte pouvant être observée dans la distribution jointe de ces variables. Pour plus de visibilité, les deux changements les plus remarquables sont soulignés dans le Tableau 2.4.

Nous remarquons également dans le Tableau 2.4 le changement de la dépendance de l'apport en eau et de la différence de température qui va de r90 Joe à la copule r90 Tawn. Cette transition n'est pas aussi remarquable que les deux premières puisque les deux copules ont des fonctions de dépendance de Pickands identiques (voir le paragraphe 1.1.3.3, page 13). Le reste des copules ne connaît pas de transition remarquable si ce n'est le changement des rotations, des valeurs prises par les paramètres ou les deux à la fois. Nous pouvons conclure que la comparaison des copules retenues pour les périodes 1 et 2 permet de confirmer la possibilité de changement brusque dans le comportement des variables hydrologiques à cause du changement climatique à travers la transition dans la structure de dépendance de l'apport en eau et de la température maximale. En outre, nous pouvons rajouter que le changement climatique a affecté le comportement joint de l'envasement et de la vitesse du vent.

Nous poursuivons l'analyse des résultats du Tableau 2.4, cette fois en comparant les copules retenues dans les périodes 2 et 3, nous pouvons voir que la plupart des copules ne subissent pas de grands changements. Ce comportement peut s'expliquer par la stabilité des comportements joints de la plupart des variables étudiées au cours de la dernière décennie. Nous remarquons de nouveau la présence de la transition entre la copule de Joe et la copule de Tawn, cette fois dans la dépendance

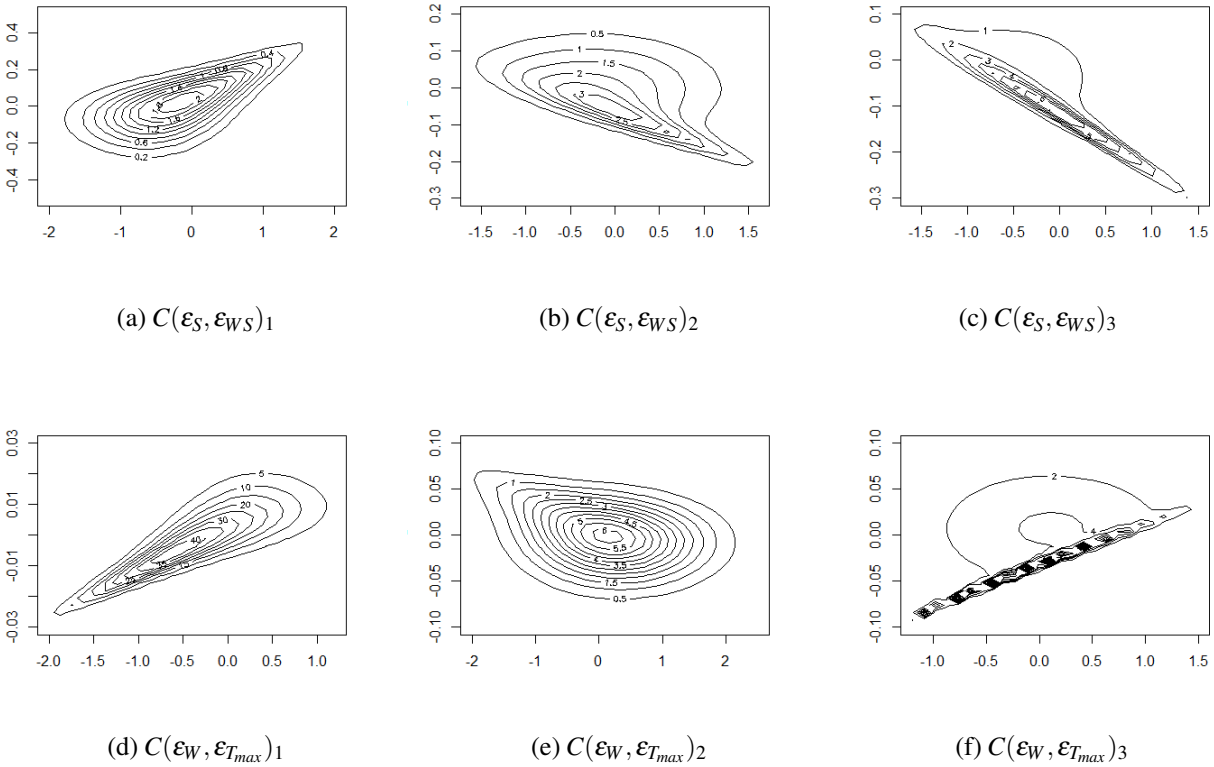


FIGURE 2.6 – *Changements remarquables des contours des densités des copules.*

entre l’apport en eau et les températures moyennes et maximales. La copule liant l’envasement et l’apport en eau connaît elle aussi un changement en passant de la copule r270 Gumbel à la copule de survie de Joe. Mais en tenant compte de la relation entre les copules de Joe et Tawn qui est l’équivalent asymétrique de la copule de Gumbel (voir la Figure 1.2, page 14), ce changement ne peut pas être considéré comme important. En général, la plupart des comportements joints des périodes 2 et 3 sont représentés par la copule de Tawn. Compte tenu de la taille des données, cela peut suggérer que la copule de Tawn est le choix le plus approprié lorsqu’il s’agit de données de taille limitée.

Quant à la période 23, nous pouvons voir que les copules restent inchangées ou subissent l’une des transitions Tawn-Joe, Tawn-Gumbel ou Joe-Gumbel précédemment discutées. Par conséquent, aucun changement significatif du comportement joint des variables étudiées ne peut être signalé au cours de cette période.

Pour la commodité du lecteur, nous continuons la discussion les résultats du Tableau 2.4, mais

de façon sommaire, en commentant l'évolution temporelle des contours des densités des copules (voir annexe A.6, Figures A2, page 95 et A3, page 96).

2.3.6 Préviation de l'envasement

2.3.6.1 Restriction du choix des copules

Afin de construire un modèle prédictif de l'envasement, nous choisissons la copule qui nous servira pour la préviation parmi les copules du Tableau 2.4, page 55. En tenant compte du changement des structures de dépendance à la fin de la période 1, les copules de cette période seront écartées. Parmi les copules restantes, nous retenons uniquement celles entre l'envasement et les variables climatiques, qui pour rappel, ont l'avantage d'être disponibles, contrairement aux variables relatives aux barrages. Finalement, nous utilisons les résultats du test d'ajustement de Cramér-Von Mises comme distance pour choisir pour chaque période les copules les mieux ajustées aux données. Ces copules ont été mises en gras sur le Tableau 2.4. Nous remarquons qu'elles sont toutes liées à la température moyenne.

2.3.6.2 Vérification de la qualité d'ajustement par simulation

Afin de visualiser la qualité d'ajustement des trois copules résultant de l'étape précédente, nous simulons avec chacune d'elles un échantillon de $n = 1000$ nouvelles observations et les ajoutons à celles existantes. Nous allons vérifier que ces copules préservent l'homogénéité des données regroupées comme décrit ci-dessous :

D'après l'inégalité de Tchebychev (voir *Monhor* [2007]), nous avons

$$P(\xi \in B(m, r)) \geq 1 - \frac{\sum_{i=1}^n \text{Var}(\xi_i)}{r^2}, \quad (2.1)$$

où $B(m, r)$ est le disque centré en m de rayon $r > 0$, $\xi = (\xi_S, \xi_T)$ est le vecteur des résidus associés à l'envasement observé et à la température moyenne, $m = (m_1, m_2)$ le vecteur moyen des résidus (observés et simulés) de l'envasement et de la température moyenne. Nous choisissons r tel qu'au moins 75% des observations originales appartiennent au disque $B(m, r)$, et conservons les copules sous-jacentes.

Les résultats illustrés sur la Figure 2.7, montrent que plus de 75% des observations originales (en gris) se trouvent à l'intérieur du cercle $\mathcal{C}(m, r)$ (en rouge) pour nos trois copules, ce qui signifie que les trois simulations ont réussi.

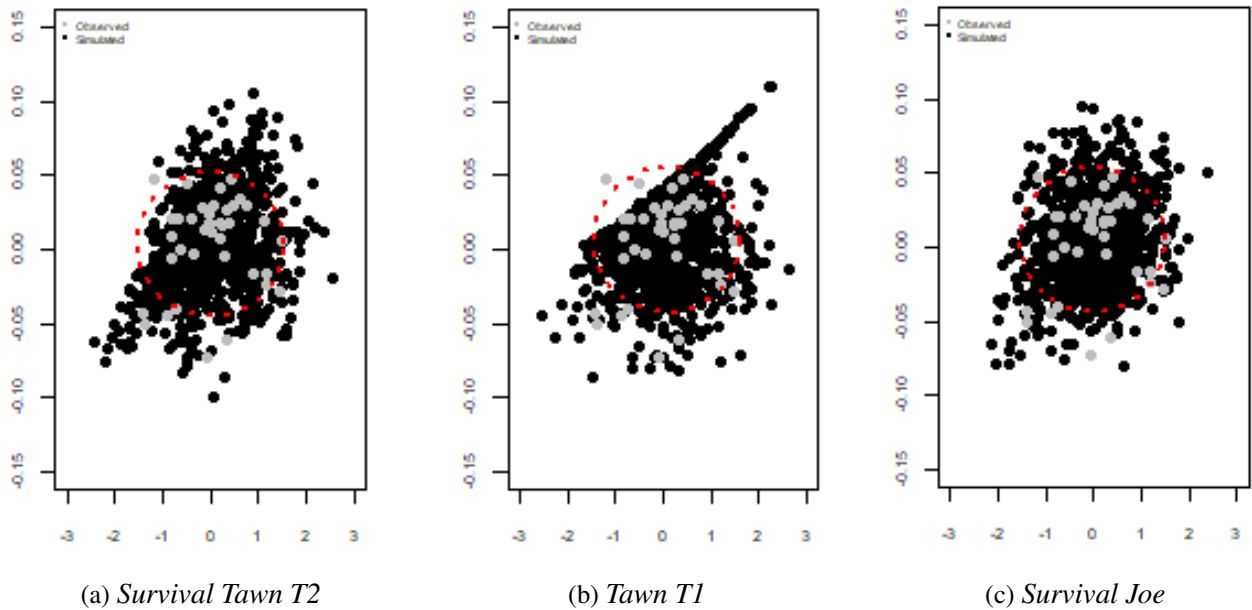


FIGURE 2.7 – Échantillons simulés de la paire des résidus $(\varepsilon_S, \varepsilon_T)$ à partir des copules ajustées, $n = 1000$.

2.3.6.3 Choix de la copule conditionnelle

Pour prévoir le résidu de l'envasement ε_S de la période 3 en fonction du résidu de la température moyenne ε_T , nous utilisons les résultats des copules conditionnelles en remplaçant U par $F(\varepsilon_T)$ et V par $F(\varepsilon_S)$ dans les équations 1.69, 1.70 et 1.71 page 30 pour les trois copules résultant de l'étape précédente. Une fois la prévision du résidu de l'envasement obtenue, nous la remplaçons dans le modèle AR(2) initial (en gras sur le Tableau 2.3, page 51) comme suit

$$\begin{aligned} \nabla(\log(\hat{S}_t)) &= \phi_1 \nabla(\log(S_{t-1})) + \phi_2 \nabla(\log(S_{t-2})) + \hat{\varepsilon}_t \\ \text{d'où} \quad \log(\hat{S}_t) &= \log(S_{t-1}) + \phi_1 (\log(S_{t-1}) - \log(S_{t-2})) + \phi_2 (\log(S_{t-2}) - \log(S_{t-3})) + \varepsilon_t \\ &= (\phi_1 + 1) \log(S_{t-1}) + (\phi_2 - \phi_1) \log(S_{t-2}) - \phi_2 \log(S_{t-3}) + \hat{\varepsilon}_t \\ &= \log(S_{t-1})^{(\phi_1+1)} + \log(S_{t-2})^{(\phi_2-\phi_1)} - \log(S_{t-3})^{\phi_2} + \hat{\varepsilon}_t \\ &= \log \left(\frac{(S_{t-1})^{(\phi_1+1)} (S_{t-2})^{(\phi_2-\phi_1)}}{(S_{t-3})^{\phi_2}} \right) + \hat{\varepsilon}_t, \\ \text{finalement} \quad \hat{S}_t &= \frac{(S_{t-1})^{(\phi_1+1)} (S_{t-2})^{(\phi_2-\phi_1)}}{(S_{t-3})^{\phi_2}} \times \exp(\hat{\varepsilon}_t), \end{aligned} \tag{2.2}$$

où \hat{S} représente l'envasement estimé et $\hat{\varepsilon}_t$ le résidu estimé par chacune des trois copules conditionnelles.

Notons que cette prévision est différente de celle d'un modèle ARMA classique, vu que nous avons modélisé l'erreur (dont nous connaissons désormais la loi conjointe), mais nous avons besoin de prévoir la série initiale, donc un retour vers cette dernière est nécessaire.

Afin de choisir la bonne copule conditionnelle parmi les trois modèles précédents, nous utilisons des diagrammes en boîtes. Les résultats représentés dans la Figure 2.8 montrent que les modèles de la copule Tawn T1 et de la copule Survival Joe sous-estiment les quantiles extrêmes des quantités de l'envasement observé. Rappelons que nous sommes face à un problème de gestion des occurrences d'épisodes d'envasement extrêmes. Le fait de sous-estimer les valeurs extrêmes peut conduire à de fausses conclusions relativement aux politiques à mettre en oeuvre dans la gestion des ressources du barrage. Par conséquent, nous décidons de choisir le modèle de copule Survival Tawn T2 pour construire le modèle de prévision d'envasement. Notre choix est d'autant plus justifié puisque ce modèle correspond le mieux à la valeur médiane observée de l'envasement.

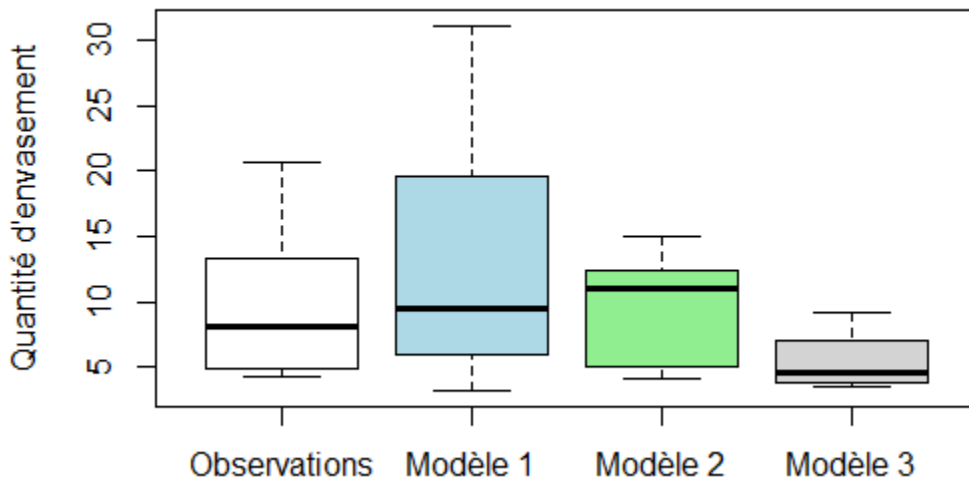


FIGURE 2.8 – Diagrammes en boîtes des modèles d'envasement. De gauche à droite : observé, copule Survival Tawn T2, copule Survival Tawn T1 et copule Survival Joe.

2.3.6.4 Résultats de la prévision

Pour expliquer l'effet de la variation de température sur l'envasement, nous simulons quatre scénarios différents de changement de température à court terme. Nous utilisons le modèle de la copule Survival Tawn T2 conditionnelle, résultant de l'étape précédente, pour prévoir le comportement de l'envasement. Nous pouvons voir sur la Figure 2.9 et le Tableau 2.5 qu'une augmentation

considérable de 2 °C par décennie de la température estivale moyenne a un effet notable sur la quantité observée de l'envasement. En revanche, l'évolution de l'envasement est modérément sensible aux variations de température moyennes à faibles (entre 0,5 et 1,5 °C par décennie). La diminution remarquable de la quantité d'envasement peut s'expliquer par la corrélation négative entre la température et les précipitations et la corrélation positive entre les précipitations et la quantité d'envasement. Un réchauffement plus prononcé conduit à la rareté des précipitations et donc à une quantité moindre d'envasement.

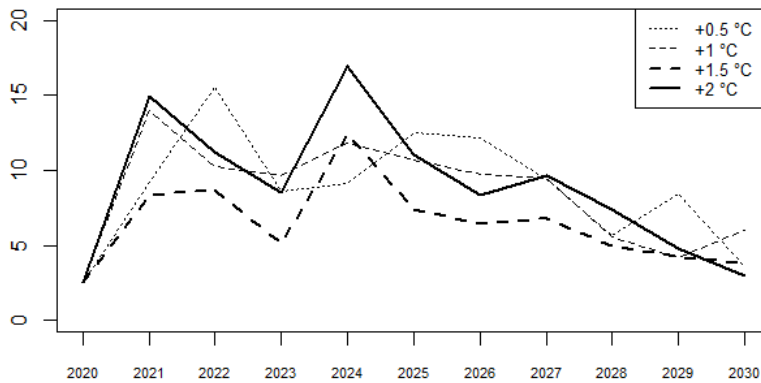


TABLE 2.5 – Valeurs prévues de l'envasement.

Année	+0,5°C	+1°C	+1,5°C	+2°C
2021	9,18	14,00	8,39	14,97
2022	15,54	10,24	8,73	11,21
2023	8,60	9,63	5,16	8,54
2024	9,17	11,79	12,43	16,98
2025	12,49	10,71	7,38	11,07
2026	12,15	9,75	6,44	8,33
2027	9,36	9,47	6,83	9,69
2028	5,63	5,61	4,96	7,37
2029	8,43	4,19	4,26	4,77
2030	3,61	5,98	3,79	2,99

FIGURE 2.9 – Prédiction de l'envasement (en Mm³) en fonction du changement annuel de la température moyenne d'été.

Soulignons que ce graphique peut ne pas être conforme à la réalité, vu que la majorité des données a été simulée, mais les bons outils ont été utilisés pour obtenir un bon modèle et cette méthodologie est complètement reproductible dans le cas où des données complètes sont disponibles.

Conclusion

Dans cette étude, nous avons confirmé grâce aux copules et certaines techniques d'ajustement le changement global proclamé des variables hydrologiques à travers une rupture structurelle de la copule reliant les paires (apport en eau, température) et (envasement, vitesse du vent) suite au changement hydrologique de 1979. Nous avons également obtenu la prévision de l'envasement de la prochaine décennie en utilisant la copule Survival Tawn T2 suivant différents scénarios de réchauffement climatique allant de $0,5\text{ }^{\circ}\text{C}$ à $2\text{ }^{\circ}\text{C}$. Le modèle de cette copule semble être très sensible à une augmentation extrême de $2\text{ }^{\circ}\text{C}$ de la température moyenne estivale. Ce résultat confirme la sensibilité de l'envasement au changement climatique et plus particulièrement aux variations importantes de température. Nous avons aussi pu conclure que la copule de Tawn est la copule la plus appropriée pour représenter la plupart des comportements joints des variables climatiques et hydrologiques. En outre, elle peut aider à modéliser ces relations même avec des échantillons de données de taille limitée.

Les résultats de cette étude pourraient être utiles aux décideurs pour gérer la possibilité d'occurrence d'un envasement extrême à l'avenir en fonction de différents scénarios de changement climatique. Plus précisément, les résultats présentés dans la partie application peuvent aider les autorités décisionnelles régionales dans leur évaluation de la situation actuelle du barrage Mohammed-V. Cependant, d'autres informations restent à vérifier. En particulier, celle de la quantité exacte d'envasement éliminée lors des opérations de dragage, s'il y en a eu.

Sommairement, la modélisation de l'envasement avec les copules est un chantier très prometteur mais qui nécessite davantage de réflexion. D'une part, la qualité des prévisions obtenues via des données simulées pourrait être améliorée à travers l'utilisation des observations réelles et des mesures de données complètes. D'autre part, une approche multivariée qui utilise par exemple la théorie des copules-vignes pourrait aider à inclure plus de variables climatiques dans la modélisation et la prédiction de l'envasement.

CHAPITRE 3

RÉGRESSION QUANTILE PAR DES COPULES-VIGNES POUR LA GESTION D'UN BARRAGE EN CASCADES

« A basin is not a random assembly of different parts, but a geomorphological system whose parts are related to each other by a long common history. »

Nash and Sutcliffe [1970]

« Everything is related to everything else, but near things are more related than distant things. »

Tobler [1979]

« The notions of positive dependence in the multivariate case are more numerous, more complex, and their interrelationships are less well understood. »

Barlow and Proschan [1981]

Introduction

Une gestion de système de réservoirs efficace nécessite la prise de décisions appropriées concernant le rejet ou le détournement de l'eau à un moment t ou à différentes périodes de l'année. En plus d'être d'une importance cruciale, ces choix sont extrêmement complexes en raison des multiples utilisations des réservoirs pour répondre, au mieux, à des besoins divers tels que la lutte contre les inondations, la production hydroélectrique, l'irrigation, l'approvisionnement en eau potable et industrielle, la navigation et les sports nautiques.

La prise de décisions managérielles est, en pratique, dictée par des règles et des politiques opérationnelles bien définies. Ces règles se présentent souvent sous la forme de courbes précisant les mesures à prendre en fonction de l'état actuel du système. Le choix de la décision nécessite des techniques d'optimisation examinant toutes les alternatives possibles (voir *Oliveira and Loucks* [1997]). Il existe une multitude de méthodes qui peuvent être utilisées pour la gestion du contrôle des crues, par exemple, l'optimisation stochastique implicite (ISO) (voir *Young* [1967]), l'optimisation stochastique explicite (ESO) et le paramétrage-simulation-optimisation (PSO) (voir *Celeste and Billib* [2009]). Chacune de ces méthodes a ses avantages et ses limites discutés dans *Zhang et al.* [2019]. Plus récemment, la méthode de gestion des réservoirs multiples appelée méthode d'agrégation-décomposition (AGDP) a été introduite pour résoudre les problèmes de dimensionnalité dans les systèmes de réservoirs à grande échelle. Elle a fait l'objet de plusieurs études, notamment celles de *Liu et al.* [2011], *Li et al.* [2014] et *Zhang et al.* [2019] où différentes règles de gestion sont développées.

Dans ce cadre, notre approche appelée "La vigne à réseau réduit" vient pour regrouper toutes les informations disponibles sur toutes les sections d'un bassin versant. Elle permet en particulier de, 1- déterminer les parties du bassin qui causent le plus d'inondations au printemps ; 2- transférer les informations des emplacements mesurés à ceux qui ne le sont pas ; 3- quantifier les volumes minimum et maximum attendus dans un barrage ; 4- gérer les besoins en eau ou faire face aux inondations attendues ; 5- déterminer la quantité d'eau excédentaire à transférer et la quantité d'eau à rejeter.

Nous capitalisons sur l'argumentation de *Cooke et al.* [2015] pour les vignes régulières et celle de *Kraus and Czado* [2017] pour les D-vignes sur les performances de la régression quantile relativement à d'autres modèles de régression. Dans les deux cas, les modèles de vignes font des hypothèses moins restrictives sur le choix des covariables influençant la réponse et peuvent résoudre analytiquement les quantiles conditionnels. Contrairement à l'approche classique (R-vigne

basée sur le Tau de Kendall), notre méthode permet de trouver la régression quantile pour la station en aval. De plus, l'utilisation de la R-vigne à réseau réduit élimine la condition d'ordre imposée dans une D-vigne et ajoute la nouvelle contrainte d'avoir la station à estimer en aval, facilitant ainsi les hypothèses restrictives du modèle et permettant la formulation de la forme conditionnelle.

Ce chapitre se présente comme suit : Après une présentation de la région d'étude et des variables en usage, nous expliquons comment obtenir la nouvelle R-vigne dans la Section 3.2.1. Nous détaillons ensuite dans la Partie 3.3.1, les étapes d'obtention du quantile conditionnel. Nous terminons cette partie par une conclusion où nous résumons les principaux résultats et discutons les principales difficultés rencontrées auxquelles nous proposons des alternatives.

3.1 Données

3.1.1 Région d'étude

Le bassin versant de la rivière Saint-Jean (Figure 3.1) est un bassin versant inter-provincial reliant le Québec et le Nouveau-Brunswick. C'est aussi un bassin versant international reliant le Canada et les États-Unis. Il couvre une superficie d'environ 55200 km^2 , dont un peu plus de la moitié est au Nouveau-Brunswick (voir *Newton and Burrell [2016]*).

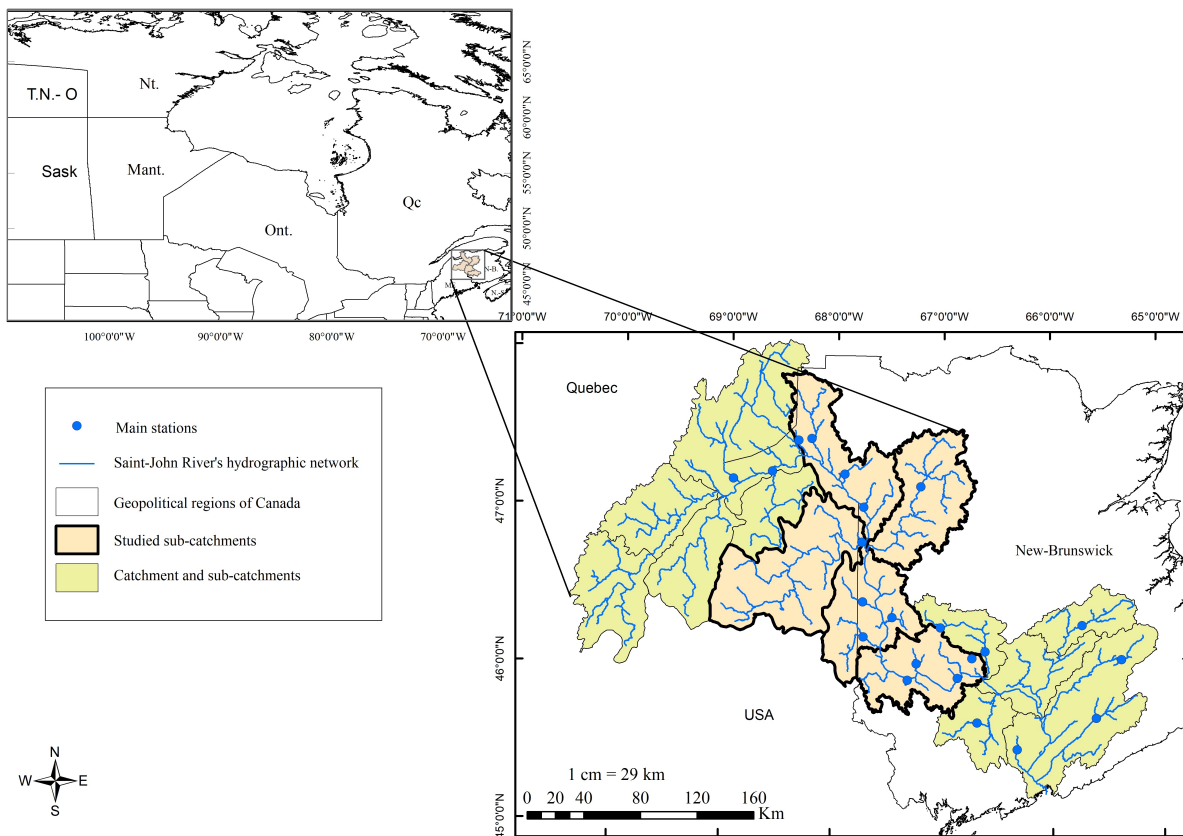


FIGURE 3.1 – *Bassin et sous-bassins de la Saint-Jean.*

La plupart des caractéristiques physiographiques du bassin du fleuve Saint-Jean sont alignées dans une direction du sud-ouest au nord-est. La zone d'amont du bassin, située à l'extrême nord-ouest, se trouve dans une petite partie des monts Notre-Dame. Cependant, une grande partie du bassin du fleuve Saint-Jean en amont de Woodstock se trouve dans le bas-plateau des Chaleurs dont l'altitude varie de 240 à 300 m. Les hautes terres du Nouveau-Brunswick s'étendent de la baie des Chaleurs au sud-ouest, puis le long de la côte de la baie de Fundy. Des parties du bas-

sin de Woodstock à Fredericton et d’Oak Point à Saint-Jean se trouvent dans les hautes terres du Nouveau-Brunswick. Les hautes terres du Nouveau-Brunswick présentent une topographie plus accidentée que le bas-plateau des Chaleurs, car les roches sédimentaires métamorphisées pénétrées par des roches éruptives sont plus résistantes à l’érosion que la roche principalement sédimentaire du bas-plateau des Chaleurs (voir *Newton and Burrell* [2016]).

La rivière Saint-Jean connaît souvent une élévation des niveaux d’eau au-dessus du seuil d’inondation dans plusieurs collectivités dont la plus remarquable est celle du printemps (avril-mai) de l’année 2008 (voir *Newton and Burrell* [2016]). Les réservoirs au Nouveau-Brunswick servent généralement plusieurs objectifs, dont le contrôle des crues, tant au niveau des affluents que du courant principal. Par conséquent, les règles de gestion devraient être en mesure de coordonner non seulement les réservoirs individuels, mais également les différentes utilisations de l’eau. Dans cette perspective, la régression quantile basée sur la R-copule-vigne est considérée comme appropriée pour un grand bassin comme celui de la rivière Saint-Jean.

Pour faire face aux épisodes d’inondations dans le bassin de Saint-Jean, la mise en place de règles de gestion efficaces est essentielle pour le décideur. C’est dans ce cadre que nous essayons d’obtenir des règles de gestion pour un système de réservoirs en limitant les pertes d’eaux des crues au niveau de cinq sections de contrôle¹ sélectionnées, sur la base de notre base de données. Nous présentons sur la Figure 3.2 la structure des réservoirs déterminée par ces stations de contrôle, choisies pour couvrir l’ensemble du bassin. Nous donnons également les codes et noms des stations dans le Tableau 3.1.

TABLE 3.1 – Codes, noms et surfaces de drainage (Km²) des stations.

Station	Code	Nom	Rivière	Superficie
S3	FX01AF002	Grand Falls	Saint John	21 900
S7	FX01AG003	Tinker	Aroostook	6 060
S8	FX01AH002	Riley Brook	Tobique	2 230
S11	FX01AJ010	Coldstream	Becaguimec stream	350
S15	FX01AK010	Mactaquac Generating Station	Saint John	39 900

Parmi les caractéristiques hydrologiques d’un épisode de crue, nous avons considéré dans notre étude le débit de pointe annuel maximal et le volume de crue correspondant². Les valeurs de ces

1. Les stations ne sont pas des sections de contrôle physiques (lac réservoir ou barrage réservoir), mais comme les stations sont presque toutes en amont des sous-bassins (exutoires), elles peuvent être considérées comme des sections de contrôle pour des réservoirs virtuels.

2. Les variables comme la durée de l’évènement pluvieux, le délai de réponse du bassin, le temps de montée de la

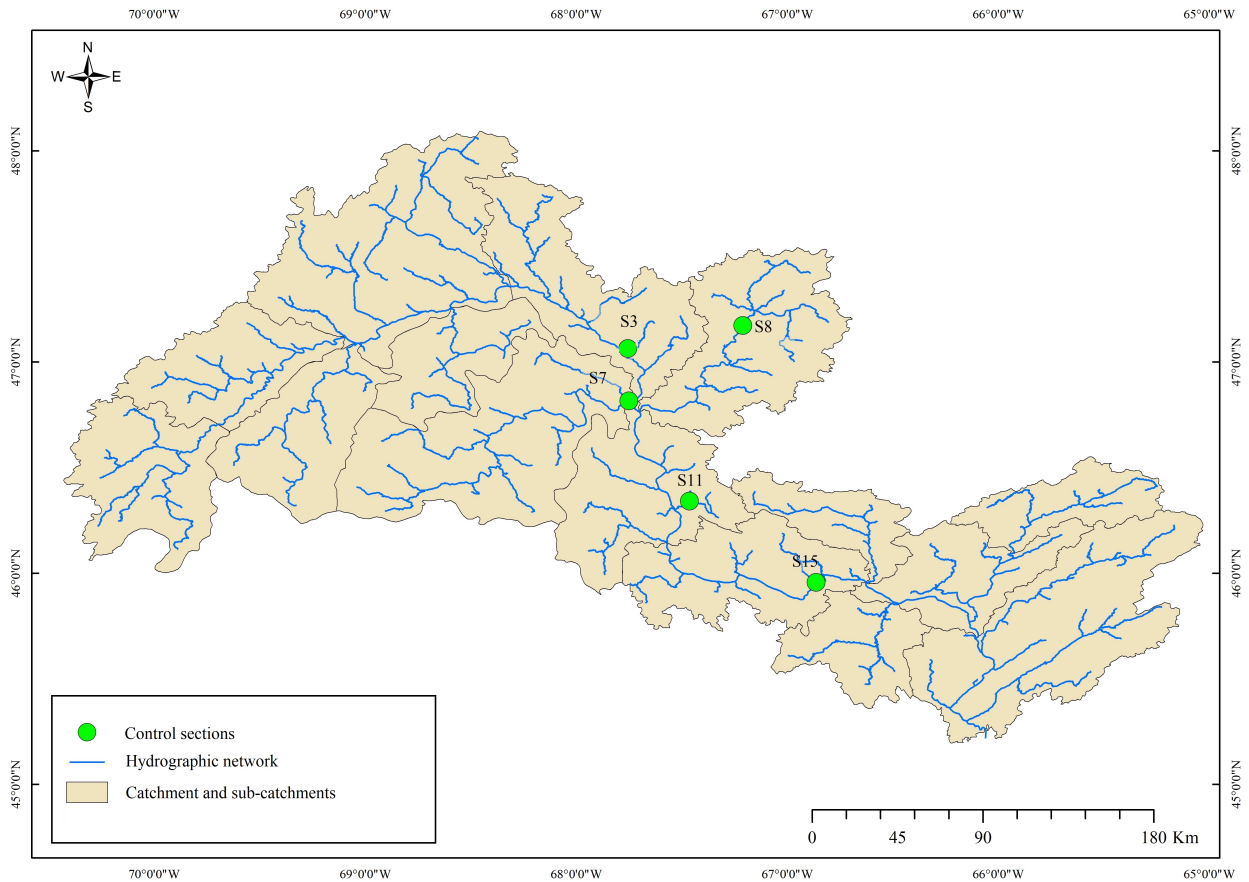


FIGURE 3.2 – Stations de contrôle sélectionnées pour illustrer la méthode R-vigne.

variables conditionnent la sévérité des crues et sont souvent utilisées dans les analyses hydrologiques. Nous avons utilisé ces deux indices pour illustrer l'utilisation de l'approche de régression quantile basée sur la copule R-vigne sur cinq sections de contrôle des crues. Nous avons alors une matrice de 10 composantes pour chacune des 42 années de la période d'observation, de 1975 à 2016 (voir Annexe B.1, Tableau B1, page 97).

Afin de faciliter la comparaison et l'interprétation, toutes les données ont été homogénéisées en divisant par la moyenne de la variable dans chaque station (voir Annexe B.1, Figure B1, page 98 pour la représentation des séries chronologiques des variables volume et débit de pointe).

3.1.2 Lois des distributions des variables

L'identification d'une distribution pouvant représenter les extrêmes observées est une étape importante dans la modélisation par R-copules-vignes. Suivant *Perreault and Bobée* [1992], sous des crue, la précipitation totale, le débit de base, les pentes de crue et de décrue ... peuvent aussi, suivant la disponibilité des données, être utilisées (voir *El-Jabi and Rousselle* [1987], *Mediero et al.* [2010] et *Pramanik et al.* [2010]).

conditions similaires, la distribution des valeurs extrêmes tend vers l'un des modèles Gumbel, Pearson II ou loi d'extremum généralisée (GEV), Weibull, Pareto et Wakeby. La distribution Gamma (Pearson type III) représente généralement les débits annuels maximaux et le volume de ruissellement annuel. La distribution de Weibull est un cas particulier de la loi GEV recommandée pour représenter les maxima. La distribution de Gumbel est un autre cas particulier de la distribution GEV lorsque son paramètre de forme $Kappa$ est nul. Ces distributions ont des queues épaisses et sont les plus pratiques pour représenter les extrêmes (voir *Perreault and Bobée [1992]*). Nous rappelons dans le Tableau 3.2 les fonctions de distribution des lois utilisées dans cette étude.

TABLE 3.2 – Description générale des fonctions de distribution utilisées dans cette étude.

Modèle de distribution	CDF	Plage des paramètres
Gamma (Pearson type 3)	$F(Q; \alpha, \beta, \zeta) = \frac{ \beta }{\Gamma\alpha} [\beta(Q - \zeta)]^{\alpha-1} e^{-\beta(Q-\zeta)}$	$Q \geq 0, \alpha, \beta > 0$
Weibull	$F(Q; \alpha, \beta) = \frac{1}{\alpha} e^{-\left \frac{Q-\alpha}{\beta}\right ^{\zeta-1}}$	$Q \geq \alpha, \beta, \alpha = 0$
Gumbel	$F(Q; \alpha, \zeta) = \frac{e^{-\frac{Q-\zeta}{\alpha}} e^{-e^{-\frac{Q-\zeta}{\alpha}}}}{\alpha}$	$-\infty < Q < \infty$
Lognormal	$F(Q; \alpha, \beta, \zeta) = \frac{-[\log(Q-\alpha)-\beta]^2}{2\zeta^2} \frac{1}{\sqrt{2\pi\zeta(Q-\alpha)}}$	$\beta < Q < \infty$

Γ = la fonction Gamma, F = CDF de Q . Source : *Langat et al. [2019]*.

Pour choisir parmi ces distributions, nous estimons par MLE leurs paramètres puis calculons les p-value du test d'Anderson-Darling (voir *Langat et al. [2019]*). Ces résultats sont regroupés dans le Tableau 3.3 où les p-values des distributions retenues sont mises en gras. Nous remarquons que la variable "Débit de pointe" suit une loi de Gumbel pour toutes les stations, tandis que la variable "Volume" suit une loi de Weibull dans les stations S7 et S11, une loi de Gamma dans les stations S8 et S15 et une loi de Gumbel dans la station S3.

TABLE 3.3 – Paramètres des distributions estimés par MLE et p-values du test Anderson-Darling.

Variable	Station	Gamma			Gumbel			Lognormale			Weibull		
		Shape	Rate	Test	Scale	Location	Test	mean	sd	Test	shape	scale	Test
Volume	3	15.23	15.23	0.59	0.21	0.88	0.66	-0.03	0.26	0.66	3.99	1.10	0.41
	7	9.58	9.58	0.42	0.31	0.86	0.25	-0.05	0.37	0.18	3.96	1.10	0.77
	8	12.26	12.26	0.89	0.24	0.86	0.81	-0.04	0.29	0.86	3.59	1.11	0.75
	11	8.16	8.16	0.76	0.32	0.84	0.57	-0.06	0.38	0.43	3.50	1.11	0.99
	15	18.47	18.47	0.67	0.19	0.89	0.66	-0.03	0.23	0.68	4.57	1.09	0.59
Débit de pointe	3	10.76	10.76	0.90	0.24	0.86	0.98	-0.05	0.30	0.98	3.06	1.11	0.30
	7	11.17	11.17	0.99	0.25	0.86	1	-0.05	0.30	0.99	3.52	1.11	0.89
	8	11.89	11.89	0.92	0.24	0.86	0.95	-0.04	0.29	0.95	3.55	1.11	0.49
	11	8.23	8.23	0.69	0.28	0.84	0.86	-0.06	0.35	0.84	2.91	1.12	0.34
	15	14.27	14.27	0.97	0.22	0.87	0.97	-0.04	0.27	0.98	3.84	1.10	0.62

Les distributions retenues et leurs résultats du test d'indépendance de Ljung-Box sont repré-

sentés dans le Tableau 3.4. Le résultat montre que les données de toutes les stations pour les deux variables ne montrent pas de corrélation en série.

TABLE 3.4 – *Distributions marginales et résultats du test d'indépendance.*

Variable	Station	Distribution	Paramètre 1	Paramètre 2	Test de Box-Ljung
Volume	3	Gumbel	0.2054731	0.8775550	0.5819
	7	Weibull	3.9564690	1.1000470	0.2502
	8	Gamma	12.2617800	12.2617400	0.6783
	11	Weibull	3.4985570	1.1118090	0.8640
	15	Gamma	18.4734700	18.4733100	0.7780
Débit de pointe	3	Gumbel	0.2391086	0.8574496	0.6619
	7	Gumbel	0.2508672	0.8574092	0.2957
	8	Gumbel	0.2432496	0.8625942	0.3618
	11	Gumbel	0.2801832	0.8363660	0.3650
	15	Gumbel	0.2201248	0.8738290	0.2701

3.2 Modélisation par R-copule-vigne

3.2.1 La nouvelle R-vigne

La structure du bassin de la Saint-Jean suggère trois différents types de représentations en copules-vignes représentés sur la Figure 3.3. Le choix de l'ordre des variables pour la structure de la vigne est un choix physique qui repose sur les connaissances apriori sur les relations entre les stations. Nous pouvons voir que c'est la R-vigne qui préserve au mieux la structure du bassin. C'est donc la structure retenue dans cette étude.

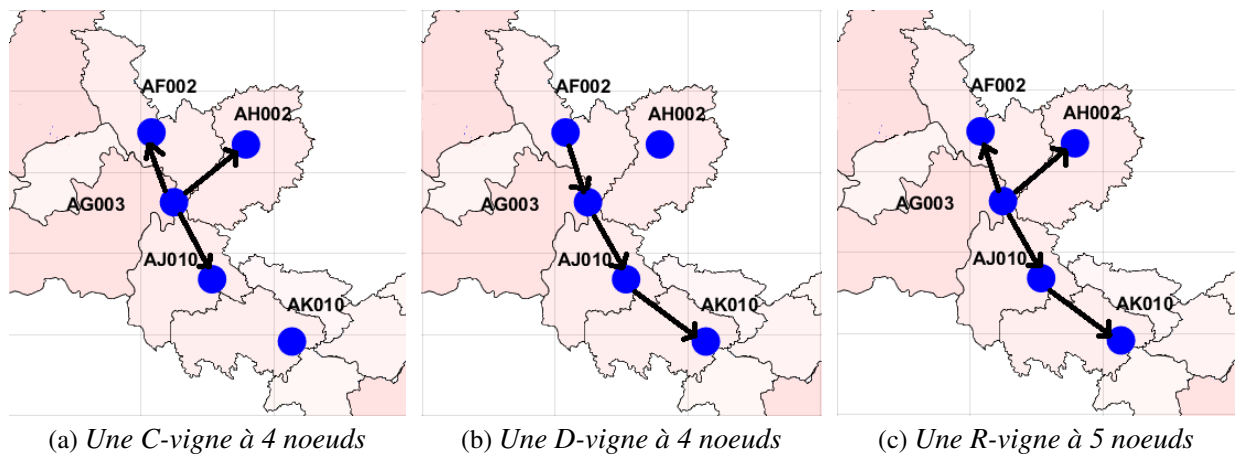


FIGURE 3.3 – Premiers arbres de trois vignes basées sur la structure du bassin. Les cercles bleus représentent les noeuds de la vigne qui sont les stations de contrôle analysées et répertoriées dans le Tableau 3.1. Les flèches noires épaisses représentent les bords de la vigne. Les lignes noires représentent les délimitations des sous-bassins versants.

La R-vigne (unique) à réseau réduit correspondant au bassin de la Saint-Jean est donnée sur la Figure 3.4.

Notons par V_1, V_2, V_3, V_4 et V_5 les variables aléatoires dans les stations 3, 7, 8, 11 et 15 respectivement.

Les données originales sont transformées en données de copule en utilisant des distributions théoriques marginales suivant l'approche PIT décrite dans le paragraphe 1.3.4, page 39. Nous effectuons maintenant une analyse par paires pour explorer la structure de dépendance entre toutes les paires de variables. Tout d'abord, nous avons déterminé les contours empiriques des données

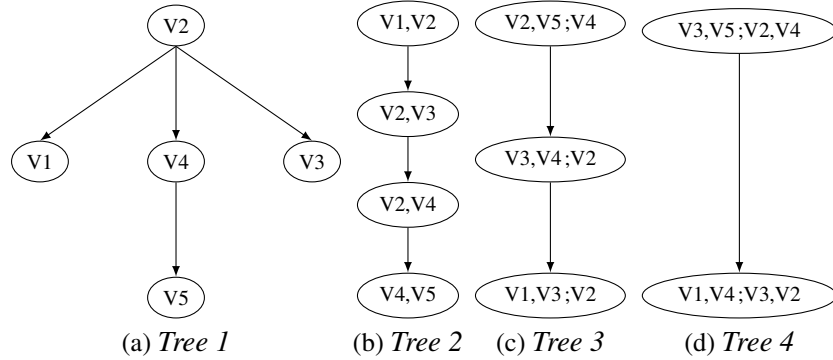


FIGURE 3.4 – La R-vigne réseau réduit.

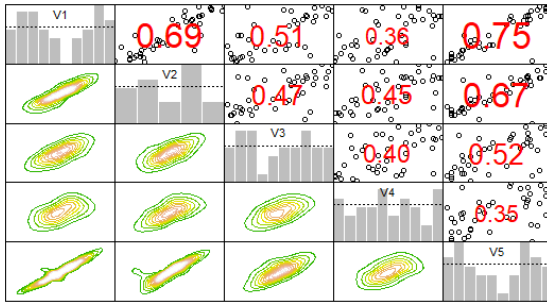
normalisées. Les résultats sont affichés sur la Figure 3.5 et montrent les dépendances non elliptiques pour les deux variables étudiées (la copule gaussienne et la copule de Student sont donc exclues). L’exploration visuelle des contours empiriques laisse suggérer qu’une copule Survival Gumbel pourrait être appropriée pour les paires (V4,V5) et (V2,V4) et une copule de Frank pour la paire (V1,V2) de la variable volume, tandis qu’une copule Gumbel pour la paire (V1,V2) et une copule Survival Gumbel pour les paires (V2,V4) et (V4,V5) de la variable débit de pointe pourraient être des choix raisonnables, respectivement.

Mis à part la méthode graphique, nous avons aussi essayé d’adapter plusieurs combinaisons de copules (Joe, Tawn et copules mixtes à deux paramètres) à nos données. Cependant, ces copules n’étaient pas appropriées pour nos données et ont donné de mauvais estimateurs. Nous avons donc limité notre choix aux copules qui correspondent le mieux à nos données : les copules de Gumbel, Clayton et Frank. Rappelons que la copule de Gumbel a un intervalle entre l’indépendance et la dépendance positive parfaite, la copule de Clayton a plus de masse dans la queue négative et la copule de Frank est une copule flexible qui ne prend aucune restriction dans son paramètre (voir Alina [2018]).

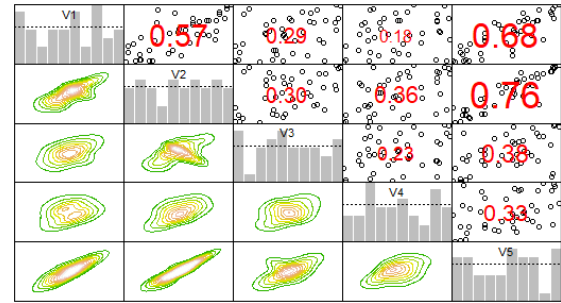
Nous donnons dans le Tableau 3.5 le résultat des copules ajustées correspondant à chaque arête de la R-vigne avec leurs paramètres. Les tracés de contours normalisés, ajustés et associés à ces copules sont présentés sur la Figure 3.6.

3.2.2 La R-vigne classique

Pour comparer la structure R-vigne proposée avec un modèle classique, nous construisons la vigne correspondant à nos données en utilisant la méthode du Maximum Spanning tree. Nous



(a) Variable volume



(b) Variable débit de pointe

FIGURE 3.5 – Contours des copules empiriques.

TABLE 3.5 – Copules bivariées et copules bivariées conditionnelles de la R-vigne à réseau réduit.

Arbre	Lien	Volume			Débit de pointe		
		Copule	Paramètre	K-Tau	Copule	Paramètre	K-Tau
1	4,5	Sur Gumbel	1.45	0.31	Sur Gumbel	1.42	0.30
	2,4	Sur Gumbel	1.65	0.39	Sur Gumbel	1.58	0.37
	2,3	Frank	5.38	0.48	Gumbel	1.50	0.34
	1,2	Frank	11.21	0.70	Gumbel	2.12	0.53
2	2,5;4	Gumbel	2.13	0.53	Gumbel	3.34	0.70
	3,4;2	Independence	-	0.00	Independence	-	0.00
	1,3;2	Independence	-	0.00	Independence	-	0.00
3	3,5;2,4	Clayton	0.67	0.25	Independence	-	0.00
	1,4;3,2	Independence	-	0.00	Independence	-	0.00
4	1,5;3,2,4	Clayton	1.18	0.37	Gumbel	1.53	0.35

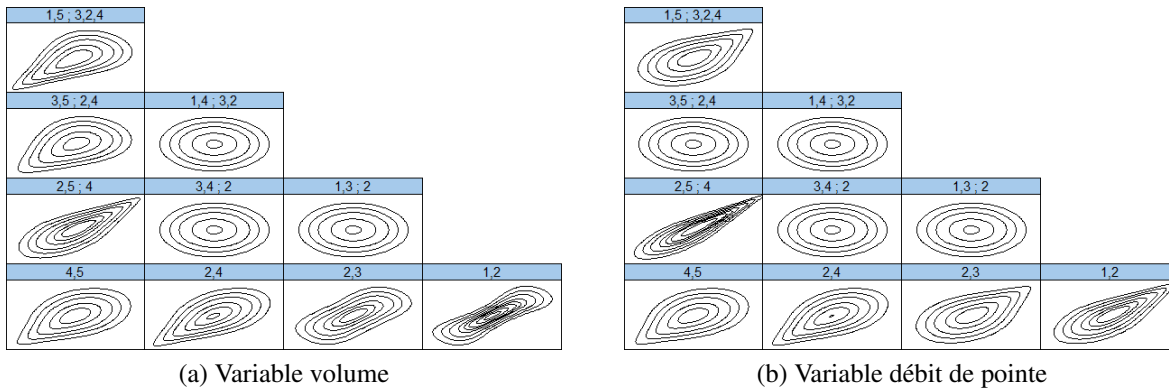


FIGURE 3.6 – Contours des copules ajustées aux données.

représentons sur la Figure 3.7 les pair-plots, les histogrammes et pour chaque paire de variables les valeurs du Tau de Kendall correspondants. Ces derniers suggèrent une structure D-vigne pour la variable volume (représentée sur la Figure 3.8)¹ et une structure R-vigne pour la variable débit de pointe (représentée sur la Figure 3.9).

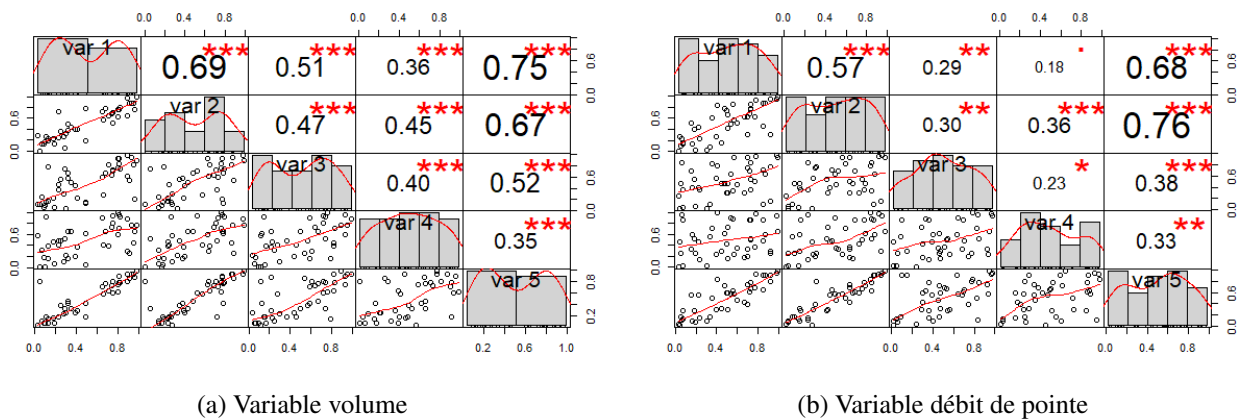


FIGURE 3.7 – Pair-plots (en bas à gauche de la diagonale), Taus de Kendall (en haut à droite de la diagonale) et histogrammes (diagonale) dans les cinq stations.

1. Le premier arbre de cette vigne est construit suivant les plus grandes valeurs du τ de Kendall comme suit : Pour $\tau = 0.75$ nous traçons le premier arête entre V1 et V5. Pour $\tau = 0.69$ nous traçons l'arête entre V1 et V2. Pour $\tau = 0.67$ nous ne pouvons pas tracer l'arête entre V2 et V5 car il ne vérifie pas la condition de proximité. Pour $\tau = 0.52$ nous traçons l'arête entre V3 et V5. Les arêtes entre V1 et V3 et entre V2 et V3 correspondant à $\tau = 0.51$ et $\tau = 0.47$ ne peuvent être tracés pour la même raison précédente. Finalement, pour $\tau = 0.45$, nous traçons l'arête entre V2 et V4. La forme finale du premier arbre représente maintenant toutes les variables.

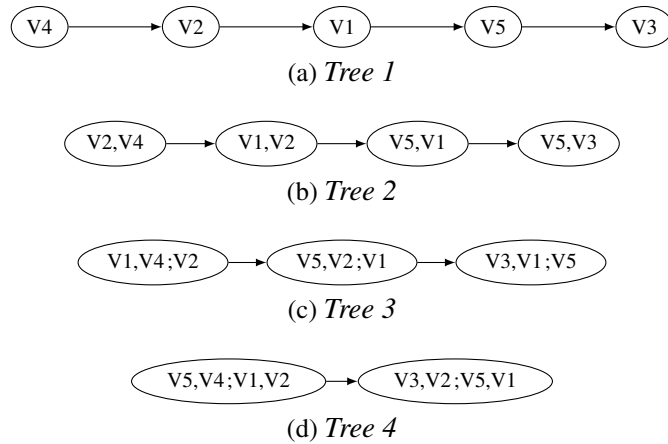


FIGURE 3.8 – Tau D-vigne de la variable volume.

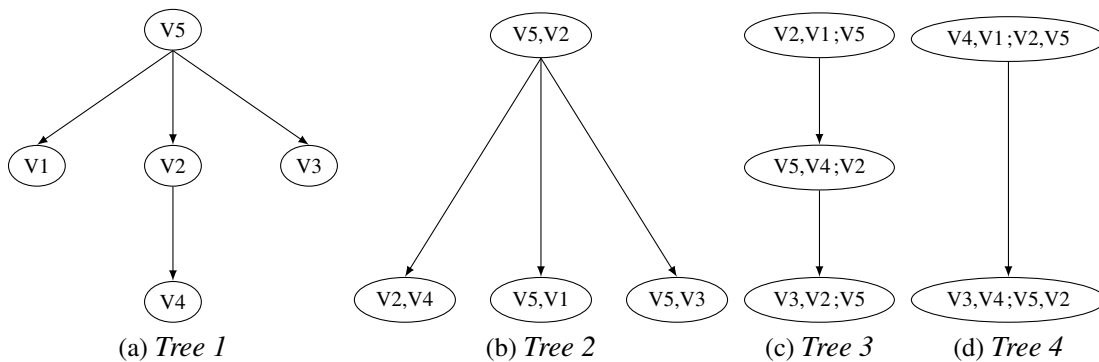


FIGURE 3.9 – Tau R-vigne de la variable débit de pointe.

3.2.3 Test de Vuong

Nous comparons l'efficacité de notre modèle pour illustrer la structure de dépendance du bassin avec celle de la D-vigne obtenue par la méthode de construction classique (que nous appellerons Tau D-Vigne) et celle d'une D-vigne avec toutes les copules de Gauss (All Gauss D-vigne) comme structure de référence. Les résultats du Tableau 3.6 montrent que le modèle Tau D-Vigne a le MLE le plus élevé, suivi par la nouvelle R-Vigne puis la All Gauss D-vigne, et ce pour les deux variables étudiées. Pour les résultats de l'AIC et du BIC, nous avons le même classement.

En plus, en examinant les résultats du test de Vuong, nous remarquons que toutes les p-values de la variable volume sont supérieures à 0.05. Nous ne pouvons donc pas rejeter l'hypothèse selon laquelle la nouvelle R-vigne est équivalente à la Tau D-Vigne et à la All Gauss D-vigne. Pour la variable débit de pointe, les p-values sont supérieures à 0.05 uniquement pour le modèle All Gauss

D-vigne, ce qui le rend équivalent à notre structure R-vigne. Par contre, les valeurs des p-values corrigées inférieures à 0.05 obtenues pour la structure Tau R-vigne combinées avec des statistiques V négatives laissent suggérer que le test favorise le modèle Tau R-vigne à la nouvelle R-vigne.

Ceci dit, même avec les résultats préférant les structures vignes obtenues de façon classique à la nouvelle R-vigne (le critère AIC à lui seul aurait suffi pour trancher (suivant *Brechmann* [2010])), nous gardons la nouvelle structure R-vigne pour faire la régression pour explorer la cohérence de ses résultats avec la position géographique des stations.

TABLE 3.6 – Comparaison des structures vignes de la variable volume.

		Volume			Pic		
		R-N vigne	K-Tau vigne	All-G vigne	R-N vigne	K-Tau vigne	All-G vigne
	Log-vraisemblance	86.17	97.2	87.98	81.1	87.13	84.22
	AIC	-158.34	-182.41	-155.96	-150.2	-166.26	-148.43
	BIC	-146.18	-171.98	-138.58	-139.78	-159.31	-131.06
	Paramètres	7	6	10	6	4	10
Statistic	Plain		-1.590321	-0.297565		-1.476475	-0.779407
	Akaike		-1.734482	0.196596		-1.966133	0.221068
	Schwarz		-1.859734	0.625941		-2.391564	1.090315
P-value	Plain		0.111763	0.766035		0.139816	0.435740
	Akaike		0.082833	0.844143		0.049283	0.825040
	Schwarz		0.062923	0.531354		0.016777	0.275574

3.3 Régression quantile

3.3.1 Expression du quantile conditionnel

L'expression de la fonction h de la vigne en réseau réduit du bassin de la Saint-Jean est détaillée comme suit. Comme dans *Aas et al.* [2009], la densité de distribution relative à une R-vigne à 5 noeuds est :

$$f_{12345} = \overbrace{f_1 \cdot f_2 \cdot f_3 \cdot f_4 \cdot f_5} \cdot \overbrace{c_{12} \cdot c_{23} \cdot c_{24} \cdot c_{45}} \cdot \overbrace{c_{13/2} \cdot c_{34/2} \cdot c_{25/4}} \cdot \overbrace{c_{14/23} \cdot c_{35/24}} \cdot \overbrace{c_{15/234}}, \quad (3.1)$$

où les termes groupés représentent respectivement les densités des distributions univariées, les densités des copules inconditionnelles de l'arbre 1, les densités des copules 1-conditionnelles de l'arbre 2, les densités des copules 2-conditionnelles de l'arbre 3 et les densités des copules 3-conditionnelles de l'arbre 4.

La distribution conditionnelle dans la station en aval dans notre cas d'étude est donnée par :

$$F(x_5/x_1, x_2, x_3, x_4) = \frac{\partial C_{15,234}(F(x_5/x_2, x_3, x_4), F(x_1/x_2, x_3, x_4))}{\partial F(x_1/x_2, x_3, x_4)} \quad (3.2)$$

Où les expressions des distributions 3-conditionnelles sont :

$$F(x_1/x_2, x_3, x_4) = \frac{\partial C_{14,23}(F(x_1/x_2, x_3), F(x_4/x_2, x_3))}{\partial F(x_4/x_2, x_3)} \quad (3.3)$$

$$F(x_5/x_2, x_3, x_4) = \frac{\partial C_{35,24}(F(x_5/x_2, x_4), F(x_3/x_2, x_4))}{\partial F(x_3/x_2, x_4)} \quad (3.4)$$

L'expression des distributions 2-conditionnelles sont :

$$F(x_1/x_2, x_3) = \frac{\partial C_{13,2}(F(x_1/x_2), F(x_3/x_2))}{\partial F(x_3/x_2)}, \quad F(x_4/x_2, x_3) = \frac{\partial C_{43,2}(F(x_4/x_2), F(x_3/x_2))}{\partial F(x_3/x_2)} \quad (3.5)$$

$$F(x_3/x_2, x_4) = \frac{\partial C_{34,2}(F(x_3/x_2), F(x_4/x_2))}{\partial F(x_4/x_2)}, \quad F(x_5/x_2, x_4) = \frac{\partial C_{52,4}(F(x_5/x_4), F(x_2/x_4))}{\partial F(x_2/x_4)} \quad (3.6)$$

Et les expressions des distributions 1-conditionnelles sont :

$$F(x_1/x_2) = \frac{\partial C_{1,2}(F(x_1), F(x_2))}{\partial F(x_2)}, \quad F(x_3/x_2) = \frac{\partial C_{2,3}(F(x_2), F(x_3))}{\partial F(x_2)} \quad (3.7)$$

$$F(x_4/x_2) = \frac{\partial C_{2,4}(F(x_2), F(x_4))}{\partial F(x_2)}, \quad F(x_2/x_4) = \frac{\partial C_{2,4}(F(x_2), F(x_4))}{\partial F(x_4)} \quad (3.8)$$

$$F(x_5/x_4) = \frac{\partial C_{4,5}(F(x_4), F(x_5))}{\partial F(x_4)}. \quad (3.9)$$

Les $F(x_i)$ correspondent aux distributions univariées.

Ensuite, d'après *Kraus and Czado* [2017], nous pouvons écrire la distribution conditionnelle de X_5 étant donné (U_1, U_2, U_3, U_4) :

$$\begin{aligned} C_{X_5/U_1, U_2, U_3, U_4}(x_5/u_1, u_2, u_3, u_4) &= h_{X_5/U_1, U_2, U_3, U_4}(C_{X_5/U_2, U_3, U_4}(x_5/u_2, u_3, u_4), C_{X_1/U_2, U_3, U_4}(x_1/u_2, u_3, u_4)) \\ &= h_{X_5/U_1, U_2, U_3, U_4}(h_{X_5/U_2, U_3, U_4}(C_{X_5/U_2, U_4}(x_5/u_2, u_4), C_{U_3/U_2, U_4}(u_3/u_2, u_4)), \\ &\quad h_{U_1/U_2, U_3, U_4}(C_{U_1/U_2, U_3}(u_1/u_2, u_3), C_{U_4/U_2, U_3}(u_4/u_2, u_3))) \\ &= h_{X_5/U_1, U_2, U_3, U_4}(h_{X_5/U_2, U_3, U_4}(h_{X_5/U_2, U_4}(C_{X_5/U_4}(x_5/u_4), C_{X_2/U_4}(x_2/u_4)), \\ &\quad h_{U_3/U_2, U_4}(C_{U_3/U_2}(u_3/u_2), C_{U_4/U_2}(u_4/u_2))), h_{U_1/U_2, U_3, U_4}(h_{U_1/U_2, U_3} \\ &\quad (C_{U_1/U_2}(u_1/u_2), C_{U_3/U_2}(u_3/u_2)), h_{U_4/U_2, U_3}(C_{U_4/U_2}(u_4/u_2), C_{U_3/U_2}(u_3/u_2)))) \\ &= h_{X_5/U_1, U_2, U_3, U_4}(h_{X_5/U_2, U_3, U_4}(h_{X_5/U_2, U_4}(h_{X_5/U_4}(x_5/u_4), h_{X_2/U_4}(x_2/u_4)), \\ &\quad h_{U_3/U_2, U_4}(h_{U_3/U_2}(u_3/u_2), h_{U_4/U_2}(u_4/u_2))), h_{U_1/U_2, U_3, U_4}(h_{U_1/U_2, U_3} \\ &\quad (h_{U_1/U_2}(u_1/u_2), h_{U_3/U_2}(u_3/u_2)), h_{U_4/U_2, U_3}(h_{U_4/U_2}(u_4/u_2), h_{U_3/U_2}(u_3/u_2))))). \end{aligned} \quad (3.10)$$

En inversant, on obtient la fonction quantile conditionnelle :

$$\begin{aligned} C_{X_5/U_1, U_2, U_3, U_4}^{-1}(\alpha/u_1, u_2, u_3, u_4) &= h_{X_5/U_4}^{-1} \left\{ h_{X_5/U_2, U_4}^{-1} \left(h_{X_5/U_2, U_3, U_4}^{-1} \left(h_{X_5/U_1, U_2, U_3, U_4}^{-1}(\alpha, h_{U_1/U_2, U_3, U_4} \right. \right. \right. \\ &\quad \left. \left. \left. (h_{U_1/U_2, U_3}(h_{U_1/U_2}(u_1/u_2), h_{U_3/U_2}(u_3/u_2)), h_{U_4/U_2, U_3}(h_{U_4/U_2} \right. \right. \right. \\ &\quad \left. \left. \left. (u_4/u_2), h_{U_3/U_2}(u_3/u_2)))) \right), h_{U_3/U_2, U_4}(h_{U_3/U_2}(u_3/u_2), h_{U_4/U_2} \right. \right. \\ &\quad \left. \left. (u_4/u_2))), h_{U_2/U_4}(u_2/u_4) \right), u_4 \right\}. \end{aligned} \quad (3.11)$$

3.3.2 Résultats de la régression

Nous montrons la contribution relative de chaque station dans l'estimation finale des quantiles de la variable volume dans la station en aval (la station 15). Ces contributions sont présentées sous forme de scénarios dans le Tableau 3.7.

Nous donnons également une comparaison avec les quantiles observés et marginaux à l'aide de l'Indice Marginal (IM) et l'Indice Observé (IO) définis comme suit :

$$IM = \frac{\hat{Q}_{100}(15)}{Qm_{100}(15)}, \quad IO = \frac{\hat{Q}_{100}(15)}{Qr_{100}(15)}, \quad (3.12)$$

où \hat{Q} , Q_m et Q_r représentent respectivement le quantile estimé par régression, le quantile marginal et le quantile observé.

TABLE 3.7 – Scénarios possibles en fonction de la gravité du quantile extrême (en 1000 m³) de la station 15.

Scénarios	S 3	S 7	S 8	S 11	S15 \hat{Q}_V	IM_V	IO_V	S15 \hat{Q}_D	IM_D	IO_D
Scenario 1	Q100	Q2	Q2	Q2	141,78	0,83	0,85	27,13	2,57	2,72
Scenario 2	Q2	Q100	Q2	Q2	145,54	0,86	0,88	27,13	2,57	2,72
Scenario 3	Q2	Q2	Q100	Q2	139,56	0,82	0,84	6,35	0,60	0,64
Scenario 4	Q2	Q2	Q2	Q100	148,81	0,87	0,90	6,65	0,63	0,67
Scenario 5	Q100	Q100	Q2	Q2	300,91	1,77	1,81	27,13	2,57	2,72
Scenario 6	Q100	Q2	Q100	Q2	145,22	0,85	0,87	27,13	2,57	2,72
Scenario 7	Q100	Q2	Q2	Q100	153,73	0,90	0,92	27,13	2,57	2,72
Scenario 8	Q2	Q100	Q100	Q2	166,32	0,98	1,00	27,13	2,57	2,72
Scenario 9	Q2	Q100	Q2	Q100	133,10	0,78	0,80	10,63	1,01	1,07
Scenario 10	Q2	Q2	Q100	Q100	153,23	0,90	0,92	6,65	0,63	0,67
Scenario 11	Q100	Q100	Q100	Q2	300,91	1,77	1,81	27,13	2,57	2,72
Scenario 12	Q100	Q2	Q100	Q100	157,98	0,93	0,95	27,13	2,57	2,72
Scenario 13	Q100	Q100	Q2	Q100	300,91	1,77	1,81	11,05	1,05	1,11
Scenario 14	Q2	Q100	Q100	Q100	155,55	0,91	0,94	10,63	1,01	1,07
Scenario 15	Q100	Q100	Q100	Q100	300,91	1,77	1,81	11,05	1,05	1,11

Dans les sous-figures 3.10a et 3.10b, nous représentons les situations où le volume de la station 15 est le plus extrême. Ceci se produit lorsque les volumes dans les stations 3 et 7 sont extrêmes simultanément (scénario 5), cette situation est extrême mais l'est moins lorsque les stations 7 et 11 sont extrêmes simultanément (scénario 8). Sur les sous-figures 3.10c et 3.10d, nous pouvons voir que lorsque le volume dans la station 7 est extrême et que le volume des autres stations est fixé sur la médiane (scénario 2), il n'y a pas d'effet notable sur la station 15. C'est également le cas lorsque les stations 3 et 8 sont extrêmes simultanément (scénario 6) ou seules comme dans les scénarios 1 et 3 représentés dans les sous-figures 3.10e et 3.10f. On peut donc en déduire que la station 7 (qui se trouve en haut de la rivière principale) a un rôle clé à jouer dans la survenue d'événements extrêmes dans la station 15 lorsqu'elle est associée à des extrêmes de la station 3, également en haut du bassin mais plus éloignée de la station 15 ou la station 8, même si elle est la plus en haut du bassin, située sur un affluent et la plus lointaine de la station 15 (voir annexe B.3 Figure B4, page 102 pour la suite des scénarios).

Pour compléter les résultats, nous avons fixé les valeurs des variables des quatre premières stations aux quantiles des périodes de retour¹ 2, 20 et 100 ans. Nous avons ensuite calculé les

1. Une période de retour est le temps entre deux occurrences d'un évènement naturel d'une intensité donnée ($n +$

quantiles conditionnels de la station 15 à partir de ces valeurs pour différentes probabilités.

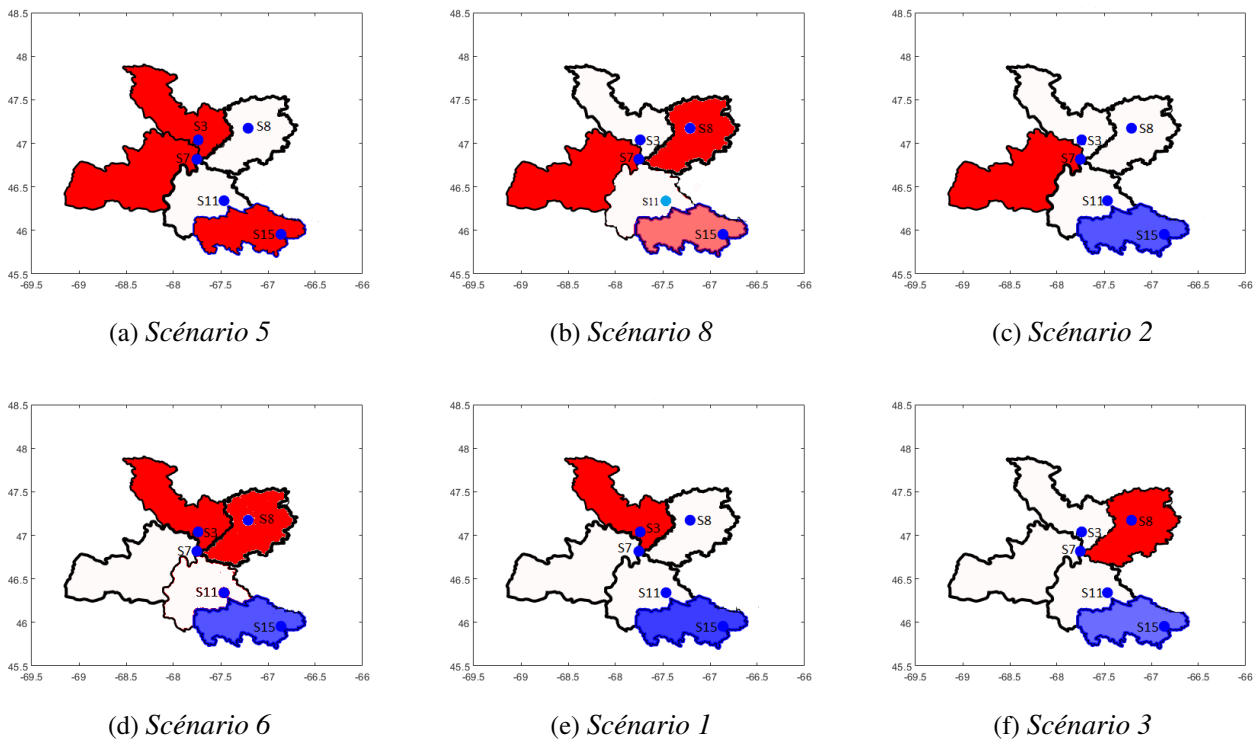


FIGURE 3.10 – Principaux scénarios de la station 15 obtenus par régression quantile. La couleur rouge signifie que les valeurs sont définies sur $Q100$ comme dans le Tableau 3.7. Le blanc signifie que les valeurs sont définies sur $Q2$. Pour la station 15, le bleu signifie que le quantile conditionnel $S15 Q100$ est inférieur à l'indice marginal, il est rose quand ils sont presque égaux et rouge lorsque $S15 Q100$ est beaucoup plus grand que l'indice marginal.

Nous présentons les résultats pour la variable volume dans la Figure 3.11 qui représente les distributions conditionnelles des événements extrêmes à la station 15 en fonction de scénarios potentiels dans les sous-bassins en amont. Il est alors possible de déduire les distributions conditionnelles des quantiles, à la station S15 suivant les niveaux enregistrés dans les quatre autres stations. Il s'agit d'une estimation basée sur le modèle multivarié et tient donc compte des incertitudes et des interactions entre les séries des extrêmes observées dans chacune des stations de mesure.

1)/ m avec n le nombre d'années et m le nombre d'évènements. L'occurrence d'un évènement pendant n années est donnée par $P(X = r) = \binom{n}{r} \times p^r \times (1 - p)^{n-r}$ (voir Hennemuth et al. [2013]).

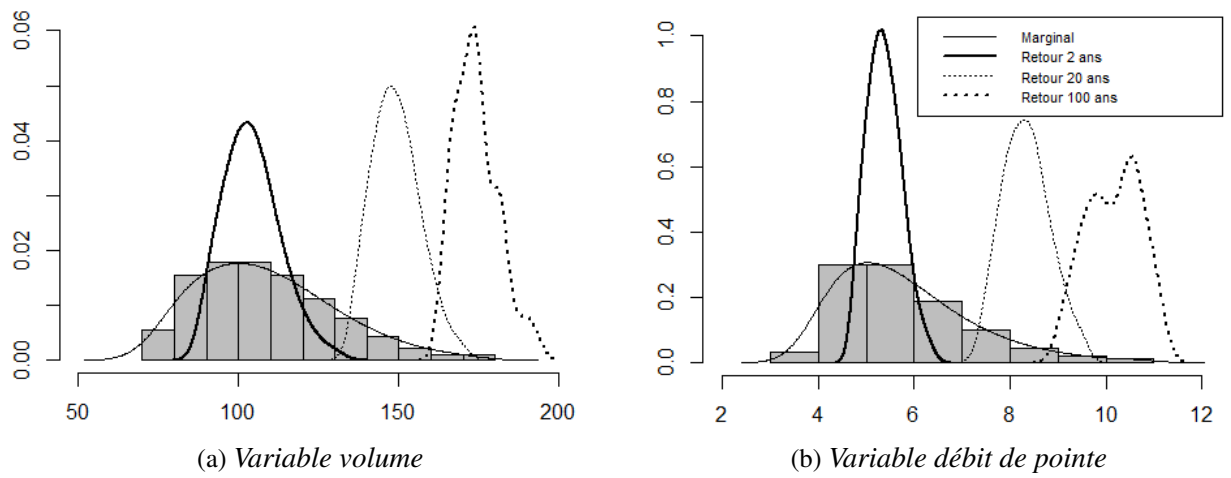


FIGURE 3.11 – Densité des quantiles conditionnels à la station 15 en fonction des variations de probabilités à covariables fixes.

3.4 Discussion

Cette étude a établi des modèles multivariés d'évaluation des risques d'inondation des réservoirs en cascade pour le contrôle des crues à partir de plusieurs points en amont. Les approches de gestion pour le contrôle des crues multi-réservoirs concernent principalement la minimisation du niveau d'eau maximal dans le barrage lui-même ou les rejets aux points de contrôle des crues en amont. Pour un grand bassin versant, comme celui de la rivière de la Saint-Jean, cette gestion est encore plus difficile. Elle doit permettre les interactions entre les sous-bassins. De plus, le décalage dans le temps et l'interdépendance des crues le long du fleuve ne doivent pas être ignorés. Le modèle copule-vigne multivarié proposé offre plus de flexibilité pour combiner toutes les sources de risque. La capacité des réservoirs existants ne peut retenir les inondations en aval lorsqu'une inondation extrême s'est produite dans le bassin supérieur du fleuve. Par conséquent, davantage de réservoirs sont nécessaires pour aider à contrôler les inondations dans certaines zones en amont. Les sous-bassins qui nécessitent plus de contrôle sont ceux connectés aux stations S3 et S7.

La rivière de la Saint-Jean possède des réservoirs de régulation saisonnière principalement utilisés pour les crues moyennes et faibles. Lorsque de grandes crues proviennent de l'amont, les réservoirs en aval pourraient ouvrir la vanne de crue prématurément, provoquant le pic de crue tôt en aval. Grâce à une meilleure compréhension du comportement des cours d'eau, de la probabilité d'inondation et des zones sensibles aux inondations en cas de débit élevé, les gestionnaires peuvent prendre des mesures, telles que le zonage des plaines inondables, pour réduire la vulnérabilité aux inondations.

L'une des extensions possibles de notre analyse serait l'utilisation de copules à valeurs extrêmes telles que Joe, Tawn ou les copules Archimédiennes à 2 paramètres, en raison de leur capacité connue à modéliser les comportements dans les queues des distributions. Cette extension pourrait être adaptée à d'autres régions hydroclimatiques. Nous pensons également qu'une extension de cette analyse pourrait concerner tous les sous-bassins versants du bassin étudié. Dans ce cas, chaque réseau de sous-bassins versants pourrait être modélisé séparément par une structure R-vigne puis tous les modèles R-vignes obtenus peuvent être regroupés en un modèle multivarié.

Conclusion

Dans ce chapitre, nous avons introduit une nouvelle approche de gestion de barrages en cascade pour combiner toutes les informations disponibles sur les différentes sections de contrôle présentes sur un bassin hydrologique. Les inondations dans un grand bassin, comme dans le cas de celui de la rivière Saint-Jean, sont un phénomène très complexe et dépendent de plusieurs paramètres temporels et spatiaux. L'approche de régression quantile basée sur la R-copule-vigne permet de réduire la dimensionnalité tout en conservant des informations essentielles sur la contribution de chaque section à la crue du bassin dans son ensemble. Une étude similaire peut être menée sur la base des prévisions climatiques pour produire des outils d'aide à la décision sur la gestion des réservoirs disponibles et aider à atténuer les effets des inondations et à protéger les zones qui y sont le plus vulnérables. Les trois principaux objectifs achevés de ce travail peuvent être résumés comme suit : (1) Conception de la distribution multivariée pour les volumes extrêmes de plusieurs réservoirs ; (2) Développement des expressions de régression quantile conditionnelle des variables hydrologiques de la station en aval ; (3) Simulation de scénarios pour les réservoirs en cascade en tenant compte à la fois des caractéristiques locales et des interactions des sous-bassins.

L'utilisation d'une structure R-vigne suggérée par le système géographique du bassin versant permet une modélisation précise de la structure de dépendance entre les stations pour une variable donnée. Le modèle proposé peut reproduire le comportement périodique des séries de débit et de débit de pointe. Le choix de la R-vigne à réseau réduit permet d'alléger les conditions imposées lors de la construction d'autres types de vigne (calcul des coefficients de corrélation et respect de l'ordre des covariables pour la D-vigne).

Les scénarios générés montrent que les conditions extrêmes dans les deux stations S3 et S7 sont déterminantes pour le sous-bassin en aval. Celles-ci indiquent la nécessité d'une amélioration de la capacité et de la gestion des volumes d'inondation au niveau de ces deux sous-bassins. En effet, la station S15 est située près de la ville de Fredericton, la capitale de la province du Nouveau-Brunswick, qui subit de fréquentes inondations.

Les quantiles prévus sont d'une importance cruciale pour les décideurs permettant une meilleure gestion des barrages existants ainsi que le développement de nouvelles structures hydrauliques. Par ailleurs, les copules-vignes pourraient être très utiles pour de nombreuses autres variables hydro-météorologiques à composante spatiale et / ou temporelle. Ces modèles permettraient également le transfert d'informations des zones mesurées vers les zones non mesurées.

CONCLUSION GÉNÉRALE ET PERSPECTIVES

Dans cette thèse, nous avons essayé de présenter les résultats de nos travaux de recherche et nos contributions dans le développement des méthodes basées sur les copules et copules-vignes en hydrologie.

Dans le Chapitre 2, nous avons pu démontrer que le changement des copules représentant les comportements joints des couples de variables hydrologiques et climatiques à travers les différentes périodes étudiées, peut être un indice sur l'avènement d'un changement de comportement des variables hydrologiques et donc, révélateur d'un changement climatique. Nous avons aussi pu obtenir grâce aux copules conditionnelles, une prévision de l'envasement à partir d'observations de variables climatiques (ici, la température). Ceci démontre que les copules peuvent être un excellent remède aux données manquantes dans le cas de variables où la collecte d'observations ne se fait pas de façon systématique, que ce soit pour des raisons techniques ou financières. Une prévision plus élargie pourrait servir de base à l'étude de l'utilité des copules dans les prévisions à moyen et long terme par exemple. Toutefois, nous jugeons utile l'extension de l'analyse des variables hydrologiques et climatiques en usage à une analyse multivariée pour garantir la prise en compte des changements simultanés et des multi-dépendances. D'autre part, et ceci a constitué l'une des limitations majeures pendant cette thèse, l'absence de données pour des périodes complètes pour certaines variables et des périodes très courtes pour d'autres (dix observations annuelles) a constitué un grand handicap, contourné, heureusement grâce à des techniques de simulation. De vraies observations auraient certainement donné des résultats plus fiables au niveau des estimations faites. Notons également que l'utilisation de données climatiques annuelles ne permet pas l'exploitation complète des fluctuations inter-annuelles et saisonnières de ces variables.

Dans le Chapitre 3, nous avons apporté une nouvelle contribution à la modélisation d'un bassin hydrologique régi par des barrages en cascade. Nous avons introduit une nouvelle approche basée sur la structure du bassin et sur l'emplacement des stations dans les sous-bassins. Cette nouvelle approche a permis l'obtention d'estimateurs des quantiles extrêmes des variables hydrologiques telles que le volume et le pic de crue, pour une station en aval conditionnellement aux valeurs prises par ces variables dans des stations se trouvant plus en amont. Cette méthodologie a un aspect très prometteur comparativement aux R-vignes classiques, où la construction des arbres est basée sur des pondérations telles que le Tau de Kendall. Des facteurs tels que la position géographique dans le bassin et la proximité entre les stations doivent être exploités au mieux pour la modélisation multivariée et la facilitation du transfert de données d'un bassin hydrologique à un autre. Nous soulignons que l'une des contraintes principales pour l'obtention de quantiles conditionnels par notre méthode est la présence de la station finale sur l'une des extrémités du 1er arbre de la R-vigne, et donc du bassin hydrologique. Cette contrainte pourrait être contournée par un "forçage" du 1er arbre de la vigne à contenir les stations d'intérêt dans les extrémités.

Une généralisation de cette méthode pour l'ensemble du bassin serait intéressante. En effet, la prise en compte des connexions entre les stations constituant le réseau hydrographique permettrait une modélisation multivariée assez complète et les quantiles conditionnels seraient certainement améliorés. D'autre part, la méthode pourrait aussi être appliquée pour une modélisation intra-station. Elle permettrait l'estimation des quantiles de variables non observées en utilisant les données des autres variables à l'intérieur d'une même station. Les stations analysées pourraient ensuite faire l'objet d'une classification suivant la nature de la vigne obtenue ou les copules retenues pour la modélisation jointe. Elles pourraient être comparées aux classifications hydrologiques et hydrographiques (position géographique, amont/aval, affluent/rivière principale, taille du sous-bassin,...) ou suivant des facteurs climatiques communs par exemple. Une autre possibilité serait de classer la sévérité des inondations suivant la nature des copules obtenues (par exemple copules extrêmes pour les inondations catastrophiques). Les affluents dans les mêmes classes pourront faire l'objet de décisions managérielles similaires. L'inclusion d'autres copules dans le choix de base, des copules à deux paramètres par exemple *Joe* [2014], est aussi envisageable vu leur flexibilité et leur capacité à représenter davantage la dépendance dans les queues des distributions.

Une autre possibilité de modélisation pourrait être la méthode entropie-copule *Singh* [2013]. En particulier, la méthode du principe d'entropie maximale (POME) *Aghakouchak* [2014], pourrait être utilisée avec les copules-vignes par exemple. Cette méthode permettrait de construire des indices d'information très utiles aux décideurs en matière de gestion des barrages, notamment dans

le cas multivarié. On pourrait aussi utiliser l'ensemble des données disponibles pour chacune des stations étudiées, dès la 1ère année d'observation, sans tenir compte de leur concomitance. L'utilisation de l'information complète certes, améliorerait l'estimation des quantiles.

D'autres questions théoriques relatives aux copules restent encore à explorer, par exemple : 1- Adaptation des copules au traitement de différents types de données (données de Panel, données censurées, données discrètes ou données mixtes); 2- L'amélioration de l'utilisation des copules-vignes dans le cas des Big data (coût de calcul, performance, ...); 3- Le développement d'autres méthodes d'estimation des paramètres des copules, autres que le MLE; 4- La comparaison avec d'autres méthodes d'analyse multivariée telles que la régression multiple ou l'ACP; 5- Développement de tests d'ajustement des copules en utilisant l'entropie.

ANNEXES

A Suite des résultats du Chapitre 2

A.1 Tests d'indépendance

A.1.1 Test de Ljung-Box

(Ljung and Box [1978]) Le test de Ljung-Box est l'un des tests les plus utilisés dans la modélisation des processus moyenne mobile intégrée auto-régressive (ARIMA). Il ne peut être directement appliqué à la série initiale, mais plutôt aux résidus du modèle ARIMA ajusté. L'hypothèse testée est que les résidus du modèle ARIMA n'ont pas d'auto-corrélation. Soit H_0 cette hypothèse et H_a l'hypothèse alternative, c-à-d les données présentent une corrélation en série. La statistique du test de Ljung-Box est définie comme suit :

$$Q = n(n+2) \sum_{k=1}^h \frac{\hat{\rho}_k^2}{n-k} \quad (13)$$

où n est la taille de l'échantillon, $\hat{\rho}_k$ est l'auto-corrélation de l'échantillon au décalage k , et h est le nombre de décalages testés. Sous H_0 , la statistique Q suit asymptotiquement une loi de $\chi_{(h)}^2$.

A.1.2 Test de Box-Pierce

(Box and Pierce [1970]) En utilisant les mêmes notations que pour le test de Ljung-Box, Le test de Box-Pierce utilise la statistique de test $Q_{BP} = n \sum_{k=1}^h \hat{\rho}_k^2$.

A.2 Tests de normalité

A.2.1 Test d'Anderson-Darling

(Anderson and Darling [1954]) Le test d'Anderson-Darling est une modification du test de Cramér-von Mises (voir Razali and Wah [2011]), il donne plus de poids aux queues de la distribution (voir Farrell and Rogers-Stewart [2006]). C'est aussi le plus puissant des tests pour les fonctions de distributions empiriques (EDF) (voir Arshad et al. [2003]). La statistique de ce test est définie comme suit :

$$W_n^2 = n \int_{-\infty}^{\infty} [F_n(x) - F^*(x)]^2 \psi(F^*(x)) dF^*(x) \quad (14)$$

où ψ est une fonction pouvant être calculée par $[F^*(x)(1 - F^*(x))]^{-1}$ (voir Razali and Wah [2011]).

A.2.2 Test de Kolmogorov-Smirnov

La statistique de test de Kolmogorov-Smirnov, est donnée par la différence absolue maximale entre la CDF empirique et la distribution spécifiée

$$D = \max |S_n(x) - P(x)|, \quad (15)$$

pour tout $-\infty < x < +\infty$ (voir *Hennemuth et al.* [2013]). L'hypothèse nulle, que $P(x)$ est la CDF correcte, est rejetée lorsque D prend une valeur élevée. Pour des tailles d'échantillons suffisamment grandes, n nous avons $P(D_n(X_1, \dots, X_n) > \frac{1,36}{\sqrt{n}}) \approx 0,05$.

A.3 Estimation des données manquantes avec MICE

MICE est une méthode d'imputation multiple utilisée pour remplacer les valeurs de données manquantes dans un ensemble de données. Pour appliquer MICE, nous avons suivi les étapes décrites dans *Azur et al.* [2011]. Nous avons créé 5 copies de notre ensemble de données de départ (Tableau A1) et itéré 50 fois les étapes ci-dessous pour chaque copie :

Étape 1 Remplacement des valeurs manquantes dans chaque variable par leurs valeurs moyennes observées.

Étape 2 Retour aux données de départ de la première variable (avant le remplacement). De cette façon, la copie de données actuelle contient des valeurs manquantes pour la première variable, mais pas pour le reste des variables.

Étape 3 Régression de la première variable sur le reste des variables via un modèle de régression linéaire ; pour pouvoir adapter le modèle à la copie de données actuelle, nous supprimons tous les enregistrements dont la première variable est manquante pendant le processus d'ajustement du modèle. Dans ce modèle, la première variable est la variable dépendante et le reste des variables sont les variables indépendantes.

Étape 4 Utilisation du modèle de régression ajusté de l'étape précédente pour prédire les valeurs manquantes de la première variable. (Lorsque la première variable sera utilisée ultérieurement comme variable indépendante dans les modèles de régression pour d'autres variables, les valeurs observées de la première variable et ces valeurs prédites seront utilisées.)

Étape 5 Répétition des étapes 2 à 4 séparément pour chaque variable pour laquelle des données

manquent (de la deuxième à la dernière variable).

Une fois que nous atteignons le cycle final (après 50 itérations), nous conservons les valeurs imputées, c'est-à-dire un ensemble de données où toutes les valeurs manquantes ont été remplacées par des valeurs de données imputées obtenues via une procédure itérative.

TABLE A1 – Données de départ.

Année	W10	W11	W12	P10	P11	P12	Tmoy7	Tmoy8	Tmoy9	Tmax7	Tmax8	Tmax9	Tdif7	Tdif8	Tdif9	WS10	WS11	WS12	SA _n	
1968	13.20	52.80	117.30	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
1969	89.70	95.60	85.40	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
1970	27.50	17.90	23.00	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
1971	30.80	80.10	52.50	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
1972	65.60	66.40	44.50	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
1973	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
1974	7.23	8.95	8.73	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
1975	NA	7.43	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	16.25
1976	28.92	27.25	75.16	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
1977	70.93	25.48	34.29	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
1978	NA	11.12	22.83	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
1979	285.10	58.02	11.66	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
1980	26.49	39.20	26.27	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
1981	17.43	5.51	25.43	0.20	0.01	1.21	23.67	24.98	22.85	30.69	32.11	29.89	12.66	12.66	12.44	1.98	1.99	3.35	NA	NA
1982	126.12	24.10	25.03	1.03	3.09	1.39	26.34	25.44	22.54	34.03	32.60	29.06	14.08	12.66	11.23	2.51	2.60	2.76	NA	NA
1983	5.89	22.99	40.83	0.23	1.55	1.45	26.31	25.12	24.72	34.13	32.17	32.24	14.11	12.67	13.09	2.12	2.13	2.77	NA	NA
1984	3.64	32.29	12.24	0.11	2.79	0.25	27.40	25.40	23.62	35.25	32.50	30.81	14.19	12.53	12.46	2.56	3.01	2.46	NA	NA
1985	14.82	82.56	47.21	0.24	2.59	1.41	27.07	26.98	24.61	34.75	34.86	31.71	13.80	14.30	12.42	2.03	2.50	2.59	10.64	NA
1986	130.67	140.04	14.63	2.11	3.69	0.94	25.98	27.29	23.29	33.59	34.68	29.78	13.74	13.14	11.28	1.98	2.37	2.28	NA	NA
1987	95.50	28.71	25.80	2.05	1.55	0.61	24.70	26.33	25.04	31.39	34.42	32.00	12.04	13.94	11.84	2.15	2.68	2.30	NA	NA
1988	61.70	36.07	21.04	1.27	1.37	0.49	26.74	28.41	23.17	34.63	35.87	30.45	14.35	13.32	12.52	2.15	2.12	2.35	NA	NA
1989	84.90	85.44	72.77	1.60	1.95	1.36	27.46	28.18	23.32	34.86	35.07	29.56	13.50	12.30	10.56	2.16	2.87	2.81	NA	NA
1990	42.98	28.57	91.97	0.53	1.35	2.46	26.27	26.94	25.17	33.27	33.87	31.51	12.56	12.45	10.72	2.25	2.25	2.64	4.30	NA
1991	53.62	9.16	21.00	1.04	0.76	0.78	26.64	27.71	23.94	33.79	35.59	30.21	13.10	13.96	11.05	2.28	2.71	1.83	NA	NA
1992	6.118	25.30	8.70	1.31	0.99	0.92	24.70	26.74	23.49	31.81	34.48	30.62	12.94	13.91	12.29	2.67	2.29	2.61	NA	NA
1993	4.33	691.22	50.28	1.47	3.68	0.30	25.83	26.72	20.97	33.24	33.88	27.78	13.28	12.63	12.11	2.39	2.61	2.57	NA	NA
1994	113.34	20.06	8.57	2.36	1.00	0.42	29.37	28.37	22.36	36.77	35.98	29.36	13.35	13.29	12.36	2.10	2.19	2.20	14.19	NA
1995	77.28	7.61	18.70	0.92	0.41	1.62	25.65	26.61	20.71	32.74	33.56	27.22	13.08	12.20	11.79	1.95	2.57	2.88	NA	NA
1996	4.77	NA	118.56	0.72	0.36	2.93	24.92	25.36	20.67	31.88	32.31	26.86	12.64	12.49	11.03	2.25	2.59	3.20	NA	NA
1997	51.651	37.51	63.80	1.01	1.69	2.07	24.14	25.65	22.88	31.48	32.88	29.13	13.11	12.79	10.64	2.14	3.30	2.76	2.57	NA
1998	8.18	4.30	5.36	0.37	0.99	0.68	27.16	26.76	23.65	35.09	34.25	30.40	14.28	13.21	11.66	2.38	2.29	2.34	NA	NA
1999	31.47	12.24	9.28	1.64	2.21	1.49	27.04	27.84	22.87	34.53	35.03	29.63	13.67	12.74	11.69	2.53	2.31	2.74	NA	NA
2000	541.53	59.08	95.81	2.94	2.44	1.90	27.21	27.65	23.03	34.85	35.71	30.09	13.92	14.17	12.08	2.44	2.62	3.12	6.63	NA
2001	136.13	59.61	83.73	0.99	3.48	2.43	26.83	27.60	23.12	34.74	34.70	29.57	14.36	12.79	11.12	2.12	2.43	2.05	20.74	NA
2002	55.27	93.62	31.16	1.51	3.53	0.93	25.64	25.48	23.34	33.02	32.47	30.19	13.59	12.51	12.04	2.27	2.93	2.48	30.97	NA
2003	148.96	188.21	152.62	2.71	4.42	2.57	28.25	27.90	23.58	35.61	35.25	30.56	13.12	13.11	11.94	2.26	2.61	2.86	4.89	NA
2004	2.44	40.21	12.90	1.87	1.40	1.37	26.29	27.44	24.40	33.78	34.80	31.72	13.26	13.08	12.60	2.27	1.88	2.77	NA	NA
2005	28.31	25.61	11.14	1.37	1.91	0.60	27.22	26.77	22.37	34.80	34.25	28.84	13.40	13.31	11.34	1.98	2.55	2.38	NA	NA
2006	45.04	53.68	20.30	0.62	0.67	1.52	28.25	26.42	23.21	35.58	33.69	29.74	13.40	12.99	11.55	2.25	2.31	2.48	NA	NA
2007	74.46	137.88	23.74	4.63	3.00	0.75	26.96	26.87	22.75	34.82	34.32	29.56	14.34	13.30	11.53	2.25	2.32	2.09	NA	NA
2008	993.39	208.71	167.39	6.25	3.95	3.34	27.17	27.59	22.92	34.32	35.05	28.73	12.66	13.14	10.09	2.88	2.98	3.07	9.45	NA
2009	34.34	3.13	99.88	0.21	1.11	2.63	28.06	26.38	21.20	35.71	33.44	26.38	13.81	12.40	9.16	1.97	3.04	3.38	NA	NA
2010	142.76	57.54	104.26	2.27	2.14	1.29	27.25	26.95	23.12	34.43	33.80	29.78	12.84	12.21	11.50	2.32	3.05	3.10	NA	NA
2011	43.02	82.00	32.142	1.79	5.60	0.63	26.91	28.21	23.57	34.37	35.63	30.26	13.66	13.09	11.507	2.26	3.06	2.19	13.28	NA
2012	46.23	138.08	63.25	2.77	5.89	1.04	27.33	29.50	23.69	35.07	37.44	30.96	13.75	14.32	12.50	2.36	2.86	2.23	NA	NA
2013	NA	60.47	42.96	0.32	1.71	2.43	25.12	26.48	22.46	32.18	33.57	28.42	12.73	12.53	10.48	1.94	2.42	2.65	NA	NA
2014	28.58	75.46	33.47	0.88	2.73	1.94	26.23	27.43	25.01	33.52	34.98	31.51	13.53	13.45	11.25	2.13	3.00	2.61	NA	NA
2015	144.55	NA	NA	1.72	0.46	0.00	29.30	27.04	23.13	37.04	34.07	29.84	14.31	12.34	11.58	1.82	2.03	1.85	NA	NA
2016	50.34	24.44	134.00	1.11	2.05	4.91	27.09	26.84	24.10	34.60	33.85	31.06	13.41	12.53	12.02	2.15	2.79	2.26	NA	NA
2017	NA	NA	NA	0.95	1.88	1.07	27.98	27.31	23.38	36.28	34.46	30.68	15.00	12.61	12.81	2.12	2.02	2.75	NA	NA

A.4 Statistiques descriptives, corrélations avec décalages et stationnarité des données

A.4.1 Statistiques descriptives

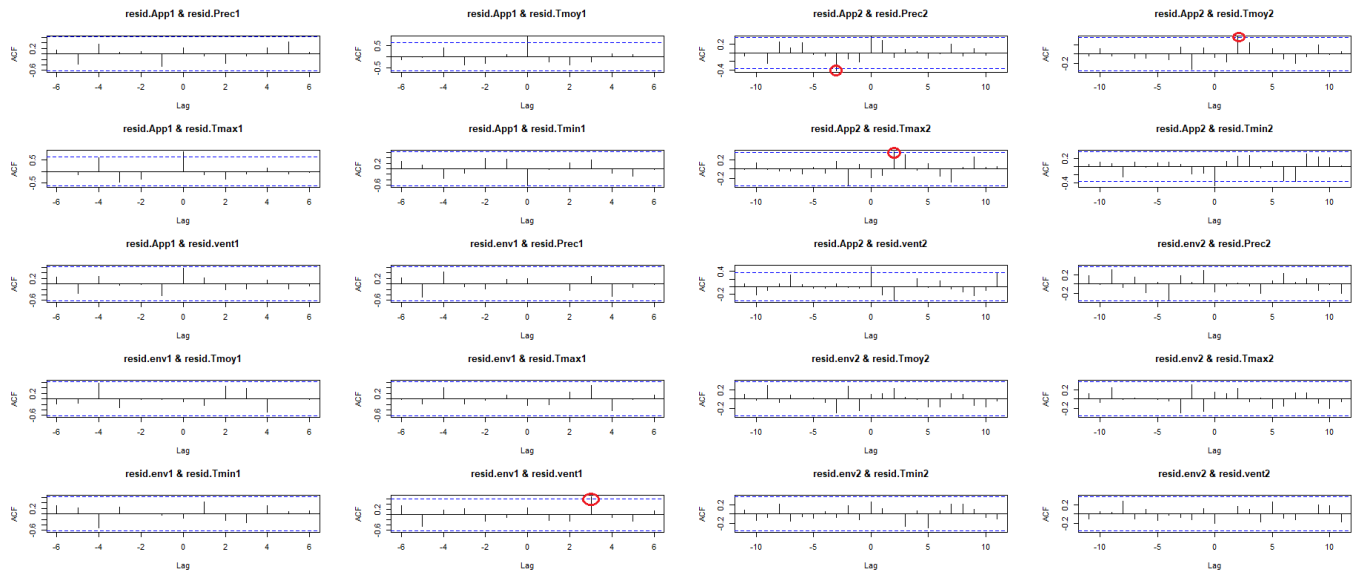
Nous présentons dans le Tableau A2 les statistiques descriptives des variables climatiques et hydrologiques.

TABLE A2 – *Statistiques descriptives des variables.*

Période	Statistique	Variables						
		W	P	\bar{T}	T_{max}	T_d	WS	S
Période 1	Min.	24.93	87.99	25.32	32.78	12.35	2.025	2.570
	1st Qu.	88.41	95.33	25.73	32.93	12.64	2.300	8.049
	Median	131.34	110.96	25.87	33.13	12.84	2.474	13.289
	Mean	134.25	130.09	25.86	33.11	12.84	2.411	12.911
	3rd Qu.	169.95	162.16	25.96	33.28	13.05	2.513	15.229
	Max.	270.70	216.42	26.54	33.58	13.43	2.732	30.974
Période 2	Min.	17.85	18.24	23.68	30.39	11.93	2.163	2.570
	1st Qu.	73.24	96.39	25.07	32.39	12.62	2.305	4.890
	Median	147.31	128.91	25.72	32.89	12.73	2.404	6.637
	Mean	231.43	145.62	25.51	32.69	12.75	2.454	9.909
	3rd Qu.	241.37	167.11	26.04	33.22	13.05	2.602	12.628
	Max.	1369.50	415.79	26.75	34.09	13.52	2.976	30.974
Période 3	Min.	65.27	67.12	24.71	31.42	11.82	1.901	4.304
	1st Qu.	137.36	121.34	25.80	32.70	12.20	2.333	4.890
	Median	157.17	169.32	26.24	33.36	12.76	2.486	9.460
	Mean	187.21	174.87	25.99	33.12	12.66	2.457	10.917
	3rd Qu.	247.58	243.02	26.26	33.69	12.77	2.580	13.289
	Max.	304.57	294.81	26.87	34.53	13.54	2.820	20.746
Période 23	Min.	17.85	18.24	23.68	30.39	11.82	1.901	2.570
	1st Qu.	87.88	98.70	25.28	32.47	12.61	2.307	4.890
	Median	150.02	133.19	25.88	32.94	12.73	2.430	6.637
	Mean	221.23	152.37	25.62	32.79	12.73	2.455	10.142
	3rd Qu.	245.35	178.05	26.14	33.40	13.04	2.595	13.289
	Max.	1369.50	415.79	26.87	34.53	13.54	2.976	30.974

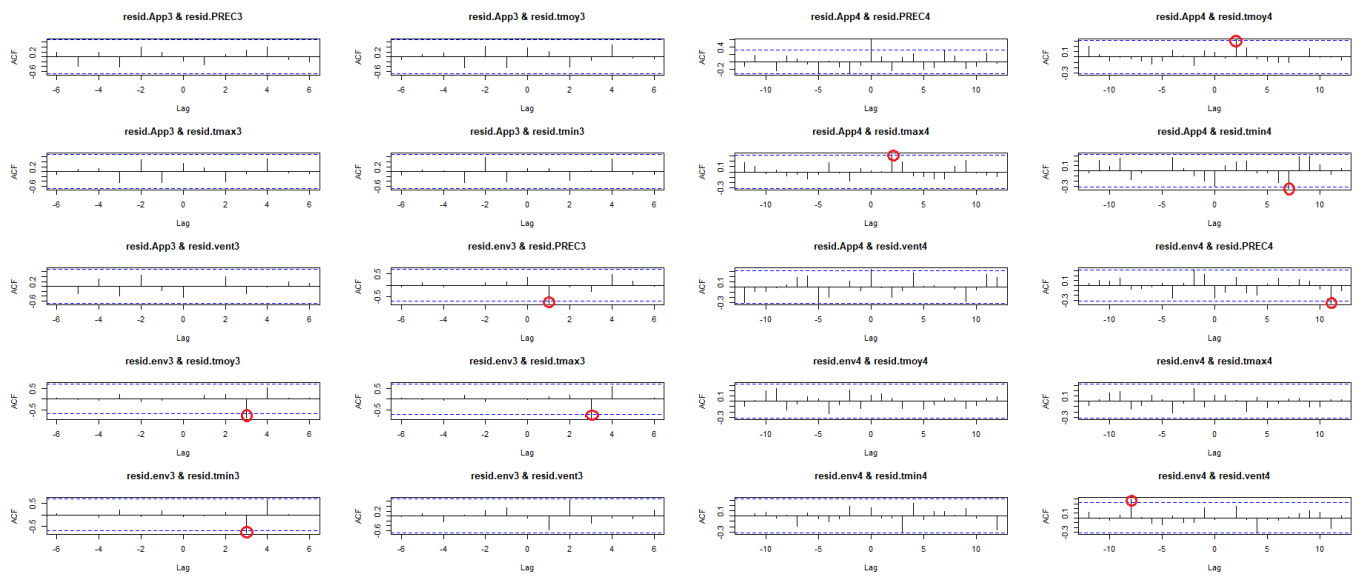
A.4.2 Corrélations avec décalages

Nous présentons dans la Figure A1 les résultats des corrélations avec décalage des séries étudiées pour les différentes périodes. Ces résultats peuvent faire l'objet d'une analyse par copules bivariées où l'une des marginales est un résidu décalé par les délais détectés k (encadrés en rouge).



(a) Période 1

(b) Période 2



(c) Période 3

(d) Période 4

FIGURE A1 – Corrélations avec décalages entre les résidus pour les différentes périodes.

A.4.3 Résultats de la stationnarité des séries

Nous présentons dans le Tableau A3 les résultats des tests de stationnarité des séries étudiées.

TABLE A3 – Résultats des tests de stationnarité des séries.

Série	Séries d'origine				Après transformation			
	D-F	p-value	KPSS	p-value	D-F	p-value	KPSS	p-value
W_t	-4.2774	< 0.01	0.0563	< 0.01	-4.5980	< 0.01	0.0508	< 0.01
P_t	-2.5099	0.3698	0.1123	< 0.01	-4.2513	< 0.01	0.0659	< 0.01
\bar{T}_t	-2.8890	0.2177	0.1155	< 0.01	-5.3139	< 0.01	0.0400	< 0.01
T_{max_t}	-3.1007	0.1328	0.1121	< 0.01	-5.3589	< 0.01	0.0404	< 0.01
T_{d_t}	-4.1003	0.0125	0.0471	< 0.01	-5.7818	< 0.01	0.0491	< 0.01
WS_t	-3.4840	0.0531	0.0497	< 0.01	-4.5320	< 0.01	0.0403	< 0.01
S_t	-4.0079	0.0164	0.0825	< 0.01	-4.4710	< 0.01	0.0385	< 0.01

A.5 Résultats MLE et AIC des copules

A.5.1 Résultats MLE

Nous représentons dans le Tableau A4 les résultats de l'estimation des paramètres des copules candidates par MLE.

TABLE A4 – Résultats de l'estimation des paramètres des copules par MLE.

Période	U_i	U_j	Clayton		Gumbel		Joe		Tawn 1			Tawn 2		
			R	Par	R	Par	R	Par	R	Par 1	Par 2	R	Par 1	Par 2
1	U_W	U_P	0	0.52	270	-1.3	180	1.52	270	-7.83	0.36	90	-7.41	0.36
	U_W	$U_{\overline{T}}$	0	2.65	180	2.22	180	3.38	180	2.28	0.53	180	2.28	0.53
	U_W	$U_{T_{max}}$	0	2.99	180	2.7	180	3.74	180	2.83	0.8	180	2.95	0.8
	U_W	U_{T_d}	270	-1.43	90	-1.68	90	-2.23	90	-3.22	0.44	270	-3.43	0.44
	U_W	U_{WS}	0	1.92	180	2.02	180	2.76	180	2.05	0.58	180	2.27	0.58
	U_S	U_P	0	0.51	180	1.24	180	1.37	0	3.32	0.4	180	5.17	0.37
	U_S	$U_{\overline{T}}$	270	-0.51	90	-1.23	90	-1.32	90	-2.0	0.16	90	-5.26	0.4
	U_S	$U_{T_{max}}$	270	-0.97	90	-1.44	90	-1.85	90	-1.7	0.31	90	-1.38	0.31
	U_S	U_{T_d}	0	0	180	1	180	1.04	180	6.27	0.31	0	5.34	0.31
	U_S	U_{WS}	180	1.43	0	1.81	0	2.24	180	2.13	0.58	0	2.19	0.58
	U_S	U_W	180	0.85	0	1.37	0	1.72	180	2.0	0.16	0	2.4	0.31
	2	U_W	U_P	0	0.65	180	1.31	180	1.51	180	4.6	0.3	0	4.8
U_W		$U_{\overline{T}}$	270	-0.29	90	-1.14	90	-1.27	90	-1.19	0.24	0	2.0	0.04
U_W		$U_{T_{max}}$	270	-0.45	90	-1.23	90	-1.39	270	-2.0	0.04	90	-1.41	0.32
U_W		U_{T_d}	270	-1.17	90	-1.63	90	-1.99	90	-3.28	0.53	270	-3.07	0.42
U_W		U_{WS}	180	1.06	0	1.54	0	1.89	0	1.64	0.48	0	2.09	0.48
U_S		U_P	90	-0.28	270	-1.12	270	-1.19	0	2.0	0.03	180	2.0	0.04
U_S		$U_{\overline{T}}$	0	0.48	180	1.22	180	1.43	0	2.12	0.3	180	2.23	0.3
U_S		$U_{T_{max}}$	0	0.42	180	1.19	180	1.38	0	2.8	0.24	180	2.31	0.29
U_S		U_{T_d}	180	0.51	0	1.26	0	1.36	0	2.0	0.2	180	2.9	0.22
U_S		U_{WS}	90	-0.55	270	-1.28	270	-1.45	90	-2.48	0.21	270	-3.47	0.34
U_S		U_W	90	-0.81	270	-1.48	270	-1.69	90	-2.05	0.48	270	-1.95	0.48
3		U_W	U_P	90	-0.62	270	-1.28	270	-1.47	90	-2.0	0.14	90	-2.0
	U_W	$U_{\overline{T}}$	0	0.99	180	1.53	180	1.77	180	9.07	0.37	180	2.0	0.3
	U_W	$U_{T_{max}}$	180	0.75	0	1.37	0	1.55	180	2.0	0.28	0	12.93	0.25
	U_W	U_{T_d}	270	-0.57	90	-1.23	90	-1.44	270	-2.0	0.2	90	-2.0	0.14
	U_W	U_{WS}	270	-1.43	90	-1.71	90	-2.22	90	-1.78	0.41	90	-2.28	0.41
	U_S	U_P	180	1.7	0	1.77	0	2.48	0	5.02	0.49	180	2.81	0.49
	U_S	$U_{\overline{T}}$	0	0.66	180	1.25	180	1.55	0	2.0	0.2	180	3.97	0.2
	U_S	$U_{T_{max}}$	90	-0.57	270	-1.24	270	-1.52	270	-3.25	0.27	90	-2.46	0.27
	U_S	U_{T_d}	90	-0.81	270	-1.45	270	-1.7	270	-4.64	0.41	90	-1.77	0.41
	U_S	U_{WS}	270	-1.58	90	-1.81	90	-2.34	90	-4.35	0.56	270	-9.08	0.49
	U_S	U_W	0	1.47	180	1.64	180	2.27	180	1.74	0.34	0	2.0	0.14
	23	U_W	U_P	180	1.6	0	1.98	0	2.44	0	2.14	0.68	0	2.32
U_W		$U_{\overline{T}}$	0	0.22	180	1.13	180	1.19	0	1.44	0.25	0	11.78	0.03
U_W		$U_{T_{max}}$	0	0.11	180	1.07	180	1.11	0	2.0	0.07	180	1.14	0.24
U_W		U_{T_d}	270	-0.52	90	-1.29	90	-1.42	90	-1.7	0.39	270	-1.43	0.39
U_W		U_{WS}	180	0.65	0	1.3	0	1.52	0	1.73	0.38	180	1.84	0.27
U_S		U_P	270	-0.35	90	-1.18	90	-1.28	90	-5.23	0.05	90	-4.68	0.09
U_S		$U_{\overline{T}}$	0	0.35	0	1.16	0	1.32	270	-2.0	0.05	180	1.77	0.26
U_S		$U_{T_{max}}$	0	0.31	180	1.12	180	1.29	270	-2.2	0.16	180	1.62	0.24
U_S		$U_{T_d}^2$	0	0.39	180	1.18	180	1.32	180	1.29	0.32	0	7.98	0.08
U_S		U_{WS}	180	0.02	270	-1.03	270	-1.05	90	-7.3	0.08	90	-19.64	0.04
U_S		U_W	90	-0.52	90	-1.3	90	-1.42	90	-1.79	0.38	270	-1.82	0.38

A.5.2 Résultats AIC

Nous représentons dans le Tableau A5 les résultats des AIC des copules pour les différentes périodes.

TABLE A5 – Résultats des AIC des copules pour les différentes périodes.

U_i	U_j	Period 1					Period 2				
		Clayton	Gumbel	Joe	Tawn 1	Tawn 2	Clayton	Gumbel	Joe	Tawn 1	Tawn 2
U_W	U_P	1.28742	1.36835	0.71134	-3.02661	-3.71303	-2.68681	-1.64616	-2.05639	-7.29391	-8.88401
U_W	$U_{\bar{T}}$	-5.77783	-3.95432	-6.20703	-0.27218	-0.27218	0.94725	0.70881	-0.00567	3.40768	1.72874
U_W	$U_{T_{max}}$	-7.97706	-7.66764	-7.90518	-4.77183	-5.04977	-0.02778	-0.37339	-0.96758	0.13868	1.94863
U_W	U_{T_d}	-1.45748	-0.97541	-1.81424	-0.97480	-0.53302	-7.59808	-8.15869	-8.21124	-12.66560	-7.62838
U_W	U_{WS}	-2.52906	-2.25557	-3.16411	0.80285	0.22297	-7.35388	-6.52865	-7.36708	-2.17916	-4.98650
U_S	U_P	1.44536	1.57355	1.56999	0.61550	-1.80554	1.15923	1.40671	1.40397	0.85287	0.22804
U_S	$U_{\bar{T}}$	1.45781	1.58679	1.68921	-3.19947	-1.04905	-0.28515	-0.35242	-1.52405	0.27851	-1.60658
U_S	$U_{T_{max}}$	0.28575	0.53828	-0.24578	2.91059	3.43763	0.09860	0.04640	-1.00722	-1.86275	-1.81933
U_S	U_{T_d}	2.00004	2.00006	1.99269	-4.15088	-1.88701	-0.74742	-0.51698	-0.24001	-14.67064	-0.87487
U_S	U_{WS}	-2.08746	-1.68839	-1.8926	1.25962	0.55490	-0.79942	-0.89272	-0.98040	0.31224	-6.09961
U_S	U_W	0.50574	0.75710	0.125821	-0.01144	2.18059	-3.49341	-4.35583	-3.81683	-2.88946	-3.01688

U_i	U_j	Period 3					Period 23				
		Clayton	Gumbel	Joe	Tawn 1	Tawn 2	Clayton	Gumbel	Joe	Tawn 1	Tawn 2
U_W	U_P	1.40865	1.51966	1.46033	0.42121	-0.07478	-19.67599	-21.07246	-19.56791	-16.17686	-17.58812
U_W	$U_{\bar{T}}$	0.30451	0.48491	0.73307	-2.11470	-4.67685	1.34772	1.04920	1.08628	2.81705	1.73973
U_W	$U_{T_{max}}$	1.07870	1.24306	1.35301	-3.06773	-0.55660	1.80331	1.68167	1.65200	-4.02820	3.76899
U_W	U_{T_d}	1.48022	1.67739	1.49808	-0.07478	0.42133	-1.47688	-2.24746	-1.96170	-1.51939	1.16501
U_W	U_{WS}	-0.59554	-0.31519	-0.77225	2.65429	1.97085	-4.12514	-3.11760	-4.31872	-1.77185	0.23087
U_S	U_P	-1.14737	-0.48748	-1.52265	-1.79834	1.39916	0.00696	0.12462	0.20881	2.32219	0.96742
U_S	$U_{\bar{T}}$	1.25868	1.59831	1.19216	-0.02277	2.26136	-0.06793	0.43692	-0.66381	-0.53812	0.59438
U_S	$U_{T_{max}}$	1.56092	1.68758	1.38437	1.53284	3.09248	0.24964	0.92508	-0.17713	1.85337	1.70476
U_S	U_{T_d}	1.16417	1.05674	1.02208	-0.57137	3.21857	-0.43941	-0.00523	-0.45425	2.43756	-1.05625
U_S	U_{WS}	-1.13084	-0.83171	-1.20585	-1.55903	-5.67003	1.99542	1.95340	1.95744	-1.11061	-0.60134
U_S	U_W	-0.69583	-0.00084	-1.02968	2.93067	0.42121	-1.77832	-2.57628	-2.06306	-1.78742	-2.20828

A.6 Analyse temporelle des copules (suite)

Sur les Figures A2 et A3 sont représentés respectivement les contours des densités des copules entre l'apport en eau et les variables climatiques, et l'engorgement et les variables climatiques pour les quatre périodes d'étude. De façon générale, nous remarquons le comportement à queue lourde caractéristique de la copule de Tawn la mieux ajustée aux paires de résidus de la période 3, qui est la plus courte des périodes étudiées. Parallèlement, les densités de la période 23, la plus longue, sont presque elliptiques et correspondent aux copules de Joe et Gumbel avec de très faibles paramètres de dépendances. Ceci laisse suggérer que les comportements joints sont devenus moins extrêmes et donc plus stables dans cette période. Une autre remarque concerne la paire $(\varepsilon_W, \varepsilon_{T_d})$, dont le contour de densité de copule semble avoir gardé le même aspect à travers le temps, la relation entre ces deux variables est donc restée globalement inchangée. Les résultats du Tableau 2.4 appuient ces remarques.

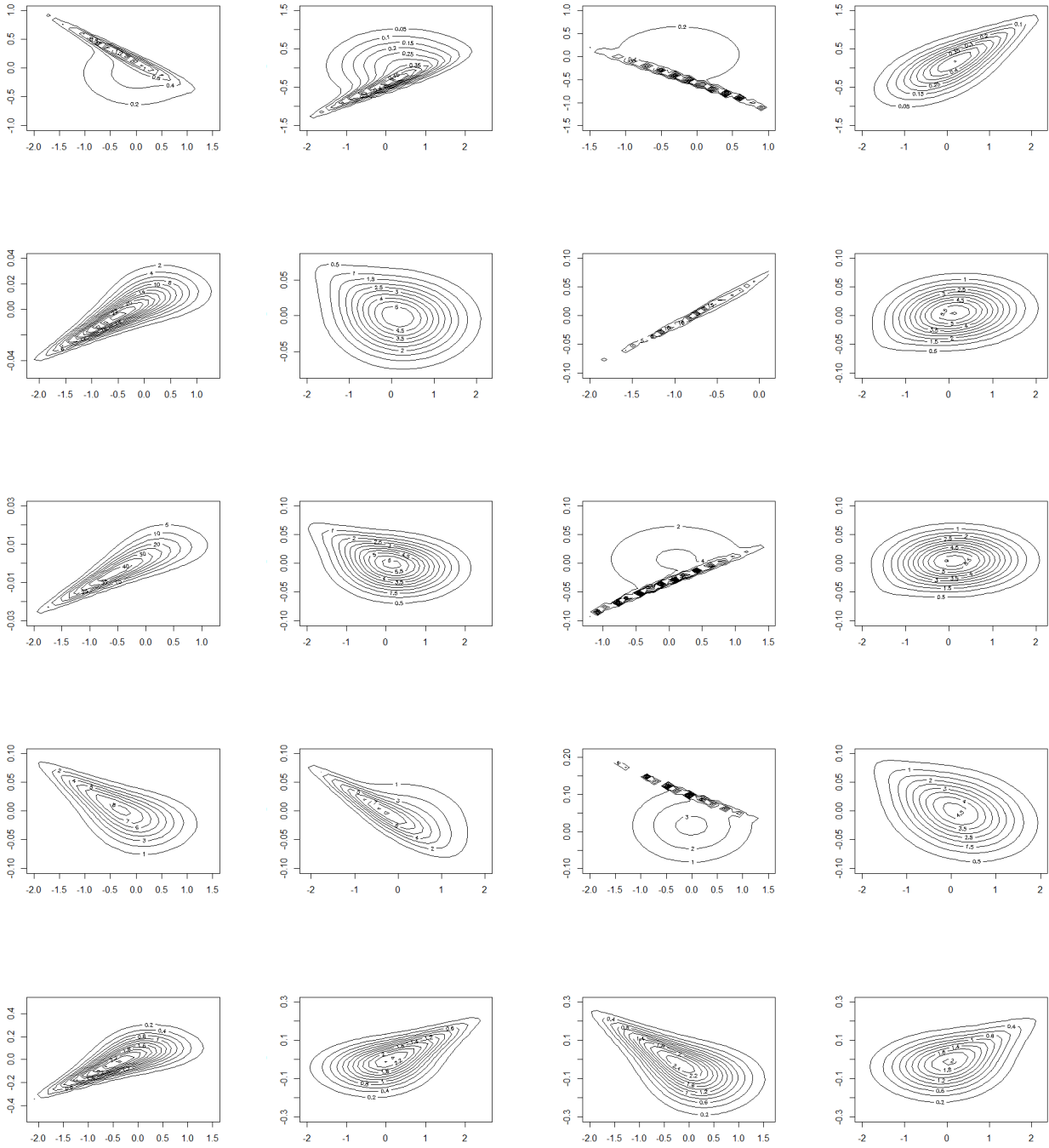


FIGURE A2 – Contours des densités des copules obtenues entre l’apport en eau et les variables climatiques. De gauche à droite : Période 1, période 2, période 3 et période 23. De haut en bas : $c(\epsilon_W, \epsilon_P)$, $c(\epsilon_W, \epsilon_{T_{max}})$, $c(\epsilon_W, \epsilon_{T_d})$ et $c(\epsilon_W, \epsilon_{WS})$.

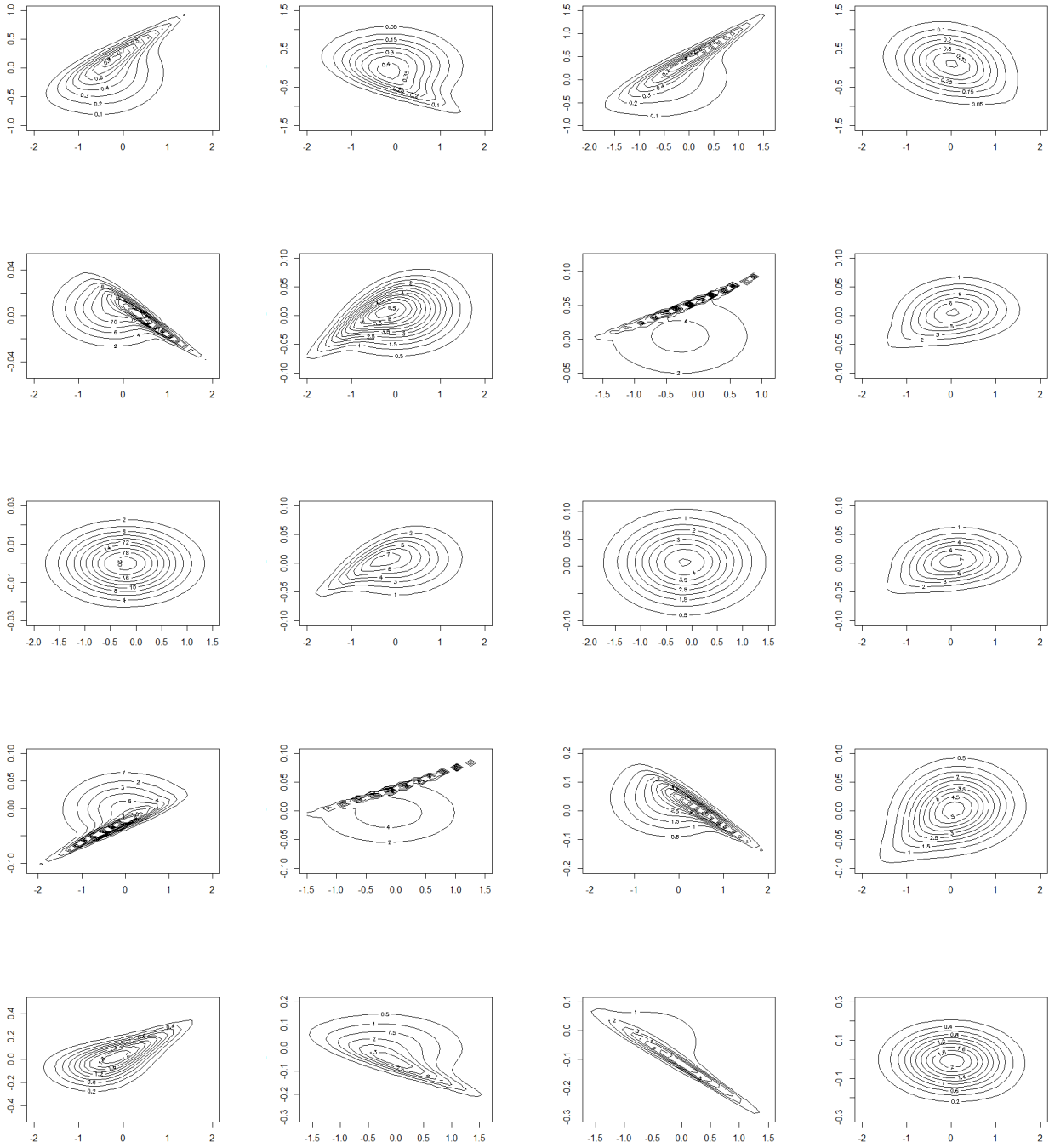


FIGURE A3 – Contours des densités des copules obtenues entre l’envasement et les variables climatiques. De gauche à droite : Période 1, période 2, période 3 et période 23. De haut en bas : $c(\varepsilon_S, \varepsilon_P)$, $c(\varepsilon_S, \varepsilon_T)$, $c(\varepsilon_S, \varepsilon_{T_{max}})$, $c(\varepsilon_S, \varepsilon_{T_d})$ et $c(\varepsilon_S, \varepsilon_{WS})$.

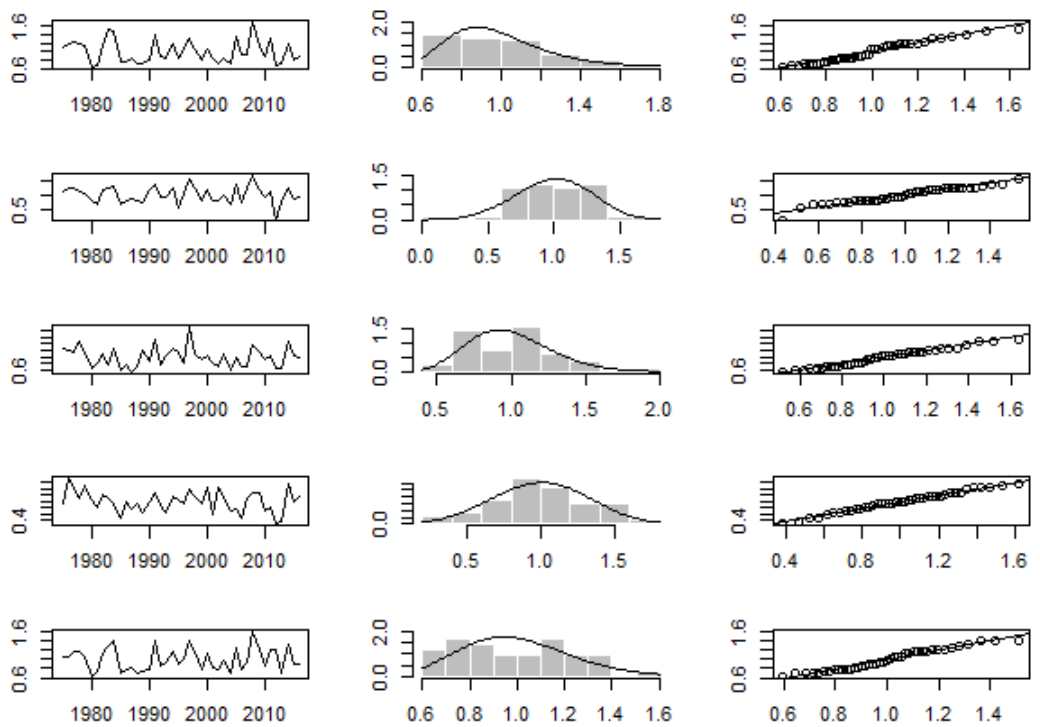
B Suite des résultats du Chapitre 3

B.1 Données

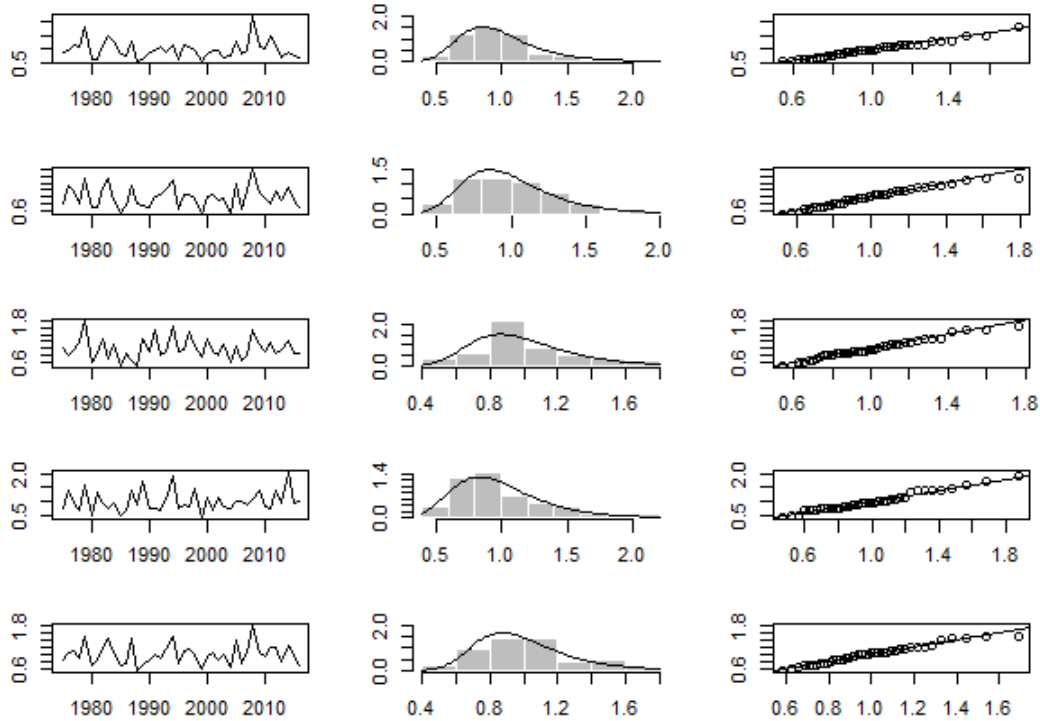
Nous représentons dans le Tableau B1 les données relatives aux variables volume et débit de pointe analysées dans le Chapitre 3. Les séries chronologiques, distributions et QQ-plots de ces mêmes variables sont représentées sur la Figure B1.

TABLE B1 – Données des variables volume et débit de pointe des stations étudiées.

Année	Volume					Pic				
	S1	S2	S3	S4	S5	S1	S2	S3	S4	S5
1975	68 405,00	17 840,00	6 216,90	973,80	112 025,00	2 990,00	776,00	247,00	55,80	4 666,00
1976	74 230,00	19 272,20	5 863,70	1 811,59	110 842,00	3 360,00	1 280,00	208,00	108,00	5 859,00
1977	75 666,50	19 744,50	5 794,90	1 541,06	123 333,50	3 860,00	1 150,00	242,00	72,20	5 995,00
1978	74 067,00	17 544,10	7 240,55	1 150,76	120 875,50	3 780,00	793,00	286,00	53,80	4 957,00
1979	70 724,00	16 281,25	5 675,85	1 583,98	110 177,00	6 250,00	1 530,00	441,00	120,00	8 441,00
1980	38 904,50	13 188,05	3 316,80	1 133,64	67 018,00	2 140,00	673,00	145,00	37,10	4 012,00
1981	42 343,00	10 796,65	3 949,00	907,66	76 593,00	2 160,00	685,00	212,00	103,00	4 590,00
1982	70 198,50	18 353,00	5 337,50	1 289,61	119 275,00	3 880,00	1 270,00	319,00	73,10	6 575,00
1983	95 719,50	19 370,50	3 793,20	1 201,38	134 822,50	5 000,00	1 510,00	174,00	59,10	7 977,00
1984	92 454,50	21 107,00	6 222,10	1 025,60	145 598,00	4 440,00	903,00	277,00	71,00	5 954,00
1985	48 129,00	10 774,50	2 966,40	546,19	72 339,50	2 780,00	538,00	127,00	36,60	3 764,00
1986	48 117,00	12 487,70	3 867,85	1 039,94	81 474,00	2 620,00	743,00	213,00	50,20	4 038,00
1987	52 321,80	13 871,15	2 698,05	789,98	87 961,00	4 400,00	1 320,00	163,00	106,00	7 926,00
1988	45 733,00	12 496,05	3 573,10	976,17	72 643,50	2 000,00	769,00	129,00	69,10	3 128,00
1989	45 836,50	11 834,75	6 036,75	720,21	79 646,00	2 080,00	745,00	317,00	129,00	3 895,00
1990	49 730,50	18 771,80	4 368,85	1 055,55	83 400,50	3 100,00	681,00	231,00	59,20	4 687,00
1991	87 730,00	21 233,50	7 581,45	1 332,59	145 424,50	3 380,00	971,00	380,00	55,90	5 592,00
1992	56 109,00	14 707,00	3 824,45	945,22	90 611,00	3 670,00	1 030,00	202,00	52,00	5 113,00
1993	53 028,00	14 589,35	5 143,35	727,11	99 116,50	3 180,00	1 180,00	229,00	79,40	6 325,00
1994	74 990,50	19 281,70	6 294,30	1 262,42	123 079,00	4 050,00	1 450,00	410,00	144,00	8 268,00
1995	52 118,00	8 646,50	5 789,15	1 089,26	92 880,00	2 060,00	636,00	229,00	56,70	4 136,00
1996	66 034,50	16 909,30	4 029,85	991,96	108 214,00	4 010,00	1 030,00	238,00	64,40	6 133,00
1997	82 334,00	24 740,85	9 436,30	1 465,49	146 690,50	3 760,00	1 030,00	370,00	62,30	6 368,00
1998	66 882,50	19 546,50	5 466,80	1 215,34	121 884,50	3 320,00	923,00	263,00	110,00	5 640,00
1999	49 820,50	12 731,50	4 690,15	984,92	82 548,50	1 920,00	491,00	192,00	30,70	3 355,00
2000	65 572,00	18 849,55	5 144,80	1 523,53	117 058,00	2 900,00	1 000,00	312,00	87,40	5 331,00
2001	51 835,50	13 152,95	4 111,08	616,72	87 010,50	3 240,00	1 060,00	230,00	50,60	5 703,00
2002	44 841,60	13 071,50	3 501,75	1 543,68	82 693,50	3 250,00	872,00	201,00	88,80	4 671,00
2003	52 990,50	15 866,50	5 472,40	1 217,70	100 575,00	2 400,00	942,00	272,00	62,00	5 535,00
2004	43 706,50	10 823,50	3 023,35	738,34	73 272,50	2 520,00	539,00	150,00	57,10	3 726,00
2005	84 537,00	21 439,00	4 970,35	847,44	132 346,00	4 530,00	1 330,00	260,00	77,90	7 783,00
2006	59 053,00	11 704,70	3 540,50	501,05	83 499,50	2 860,00	613,00	160,00	75,30	4 164,00
2007	59 156,50	19 629,20	3 387,95	1 178,83	98 578,50	3 210,00	1 120,00	202,00	66,80	5 710,00
2008	105 531,50	26 783,50	6 773,80	1 331,36	167 226,50	7 500,00	1 790,00	382,00	81,70	10 027,00
2009	70 911,50	18 579,00	5 696,55	1 380,05	124 940,50	3 850,00	1 080,00	275,00	105,00	5 785,00
2010	55 694,50	14 872,25	4 627,90	781,60	88 227,00	3 470,00	932,00	227,00	62,90	5 187,00
2011	82 168,50	17 678,00	5 242,30	879,57	126 198,00	4 980,00	802,00	295,00	56,90	6 725,00
2012	40 682,00	2 664,55	3 335,85	317,26	126 394,50	3 870,00	1 170,00	213,00	105,00	6 659,00
2013	44 140,50	11 498,90	3 197,30	369,65	72 977,50	2 500,00	864,00	243,00	72,50	4 554,00
2014	74 808,50	19 414,65	7 186,40	1 662,34	136 713,00	3 010,00	1 260,00	299,00	153,00	7 030,00
2015	50 609,30	14 207,20	5 059,20	1 064,89	95 880,50	2 600,00	888,00	217,00	70,00	5 549,00
2016	54 317,00	15 242,95	4 812,10	1 218,14	93 126,00	2 430,00	677,00	219,00	78,60	3 836,00



(a)



(b)

FIGURE B1 – *Séries chronologiques, histogrammes avec densités de lois théoriques et QQ-plots des variables volume (a) et débit de pointe (b) des cinq stations étudiées.*

B.2 Représentation matricielle de la R-vigne

Pour développer des méthodes d'inférence pour des R-vignes, *Dißmann et al.* [2013] et *Czado* [2019] utilisent des matrices pour le stockage et la représentation de la séquence d'arbres de la R-vigne. Les indices associés à ces séquences sont stockés par ordre croissant dans une matrice triangulaire supérieure. Cette matrice, si correctement représentée, permet d'effectuer le calcul de log-vraisemblance sur l'ensemble de la R-vigne. Nous allons expliquer dans ce paragraphe les étapes de construction de la matrice relative à la R-vigne de notre étude. Rappelons tout d'abord la définition d'une matrice régulière :

Définition .1. (*Matrice de vigne régulière*) (voir *Czado* [2019]) Soit M une matrice triangulaire supérieure d'entrée $m_{i,j}$ pour $i \leq j$. Les éléments $m_{i,j}$ peuvent avoir des valeurs entre 1 et d . Une matrice M est dite *matrice de vigne régulière*, si elle satisfait les conditions suivantes :

1. $\{m_{1,i}, \dots, m_{i,i}\} \subset \{m_{1,j}, \dots, m_{j,j}\}$ pour $1 \leq i < j \leq d$ (Les entrées d'une colonne donnée sont aussi contenues dans toutes les colonnes à sa droite.)
2. $m_{i,i} \notin \{m_{1,i-1}, \dots, m_{i-1,i-1}\}$ (L'entrée diagonale d'une colonne n'apparaît dans aucune des colonnes à sa gauche.)
3. Pour $i = 3, \dots, d$ et $k = 1, \dots, i-1$ il existe (j, l) avec $j < i$ et $l < j$ tels que

$$\{m_{k,i}, \{m_{1,i}, \dots, m_{k-1,i}\}\} = \{m_{j,j}, \{m_{1,j}, \dots, m_{l,j}\}\} \text{ ou} \quad (16)$$

$$\{m_{k,i}, \{m_{1,i}, \dots, m_{k-1,i}\}\} = \{m_{l,j}, \{m_{1,j}, \dots, m_{l-1,j}, m_{j,j}\}\} \text{ (Condition de proximité.)} \quad (17)$$

En reproduisant les étapes de l'algorithme général 5.1 dans *Czado* [2019] pour le calcul de la matrice régulière, nous obtenons les étapes de construction de la matrice régulière de la R-vigne du bassin illustrées sur les Figures B2 et B3. Ci-après quelques explications.

Étape 1 Nous commençons par la construction de la dernière colonne de la matrice. Nous sélectionnons 5 dans l'arête $e_4 := 5, 1; 2, 3, 4$ et toutes les entrées qui sont dans un ensemble conditionné avec 5. Les entrées encadrées en noir représentent l'ensemble conditionné $\mathcal{C}_{e_4} = \{m_{5,5}, m_{4,5}\} = \{V5, V1\}$, et celles encadrées en rouge représentent l'ensemble conditionnant $\mathcal{D}_{e_4} = \{m_{3,5}, m_{2,5}, m_{1,5}\} = \{V3, V2, V4\}$. Les arêtes de la R-vigne correspondant à cette association sont représentés en gras.

Étape 2 Étant donné le reste des arêtes représentées dans l'étape 1, nous sélectionnons 4 dans l'arête $e_3 := 4, 1; 3, 2$ et toutes les entrées qui sont dans un ensemble conditionné avec 4. Ici, $\mathcal{C}_{e_3} = \{m_{4,4}, m_{3,4}\} = \{V4, V1\}$ et $\mathcal{D}_{e_3} = \{m_{2,4}, m_{1,4}\} = \{V3, V2\}$.

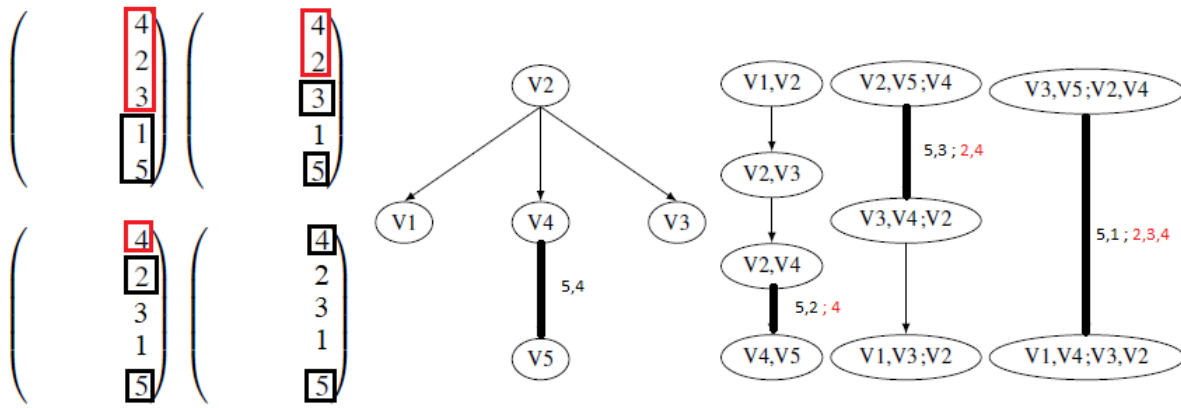
Étape 3 Pour les arêtes restantes de l'étape 2, nous sélectionnons 4 dans l'arête $e_2 := 4, 1; 3, 2$ et toutes les entrées qui sont dans un ensemble conditionné avec 4. Ici, $\mathcal{C}_{e_2} = \{m_{3,3}, m_{2,3}\} =$

$\{V3, V1\}$ et $\mathcal{D}_{e_2} = \{m_{1,3}\} = \{V2\}$.

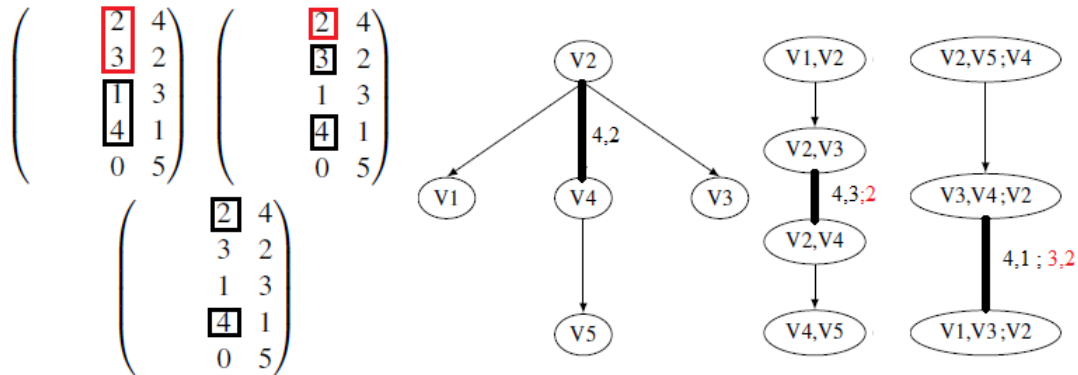
Étape 4 Pour la deuxième colonne (l'avant dernière dans l'ordre de construction), le seul arête non encore représenté est $e_1 := 2, 1$. Il est clair que $\mathcal{C}_{e_1} = \{m_{2,2}, m_{1,2}\} = \{V2, V1\}$ et $\mathcal{D}_{e_1} = \{\emptyset\}$.

Étape 5 Finalement, la matrice représentant notre R-vigne est :

$$\begin{pmatrix} 1 & 1 & 2 & 2 & 4 \\ 0 & 2 & 1 & 3 & 2 \\ 0 & 0 & 3 & 1 & 3 \\ 0 & 0 & 0 & 4 & 1 \\ 0 & 0 & 0 & 0 & 5 \end{pmatrix}$$

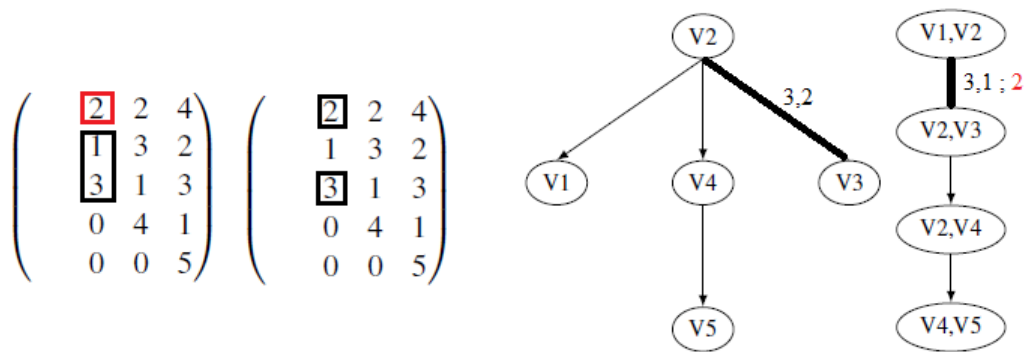


(a) Étape 1

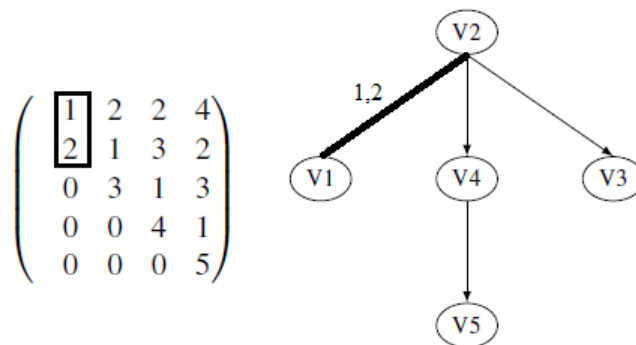


(b) Étape 2

FIGURE B2 – Étapes de construction de la matrice régulière de la R-vigne du bassin.



(a) Étape 3

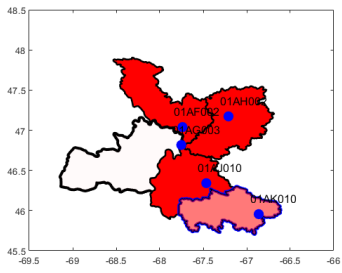


(b) Étape 4

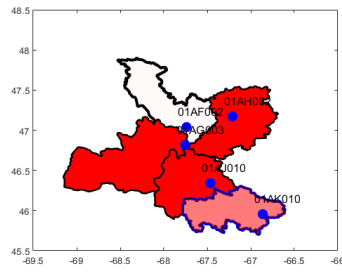
FIGURE B3 – Étapes de construction de la matrice régulière de la R-vigne du bassin (suite).

B.3 Scénarios de la régression quantile (suite)

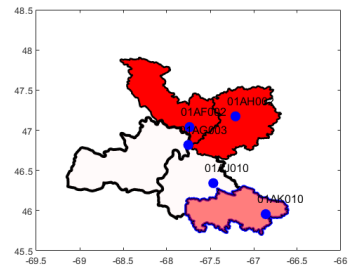
Sur la Figure B4 sont représentés le reste des scénarios de résultats de régression quantile de la station 15 pour la variable volume. La figure met en accent le rôle que joue le quantile extrême de la station 8 dans la survenue de comportements extrêmes dans la station 15. Ceci se produit par exemple lorsque S8 est extrême simultanément avec l'une des stations S3 (sous-figure B4c) ou S11 (sous-figure B4d) ou les deux (sous-figure B4a), mais jamais S8 seule (sous-figure B4e). Pour rappel, les évènements extrêmes ne survenaient pas à la station S15 lorsque S3, S11, ou les deux étaient extrêmes (voir la Figure 3.10). Par contre, les extrêmes simultanés des stations S7 et S8 (sous-figure B4f) n'ont pas d'effet notable sur la station S15. La station S8, quoique couvrant une surface de drainage relativement petite, est donc à surveiller de près, surtout quand des extrêmes sont observés dans les station S3 et S11 pour empêcher des scénarios catastrophiques en aval.



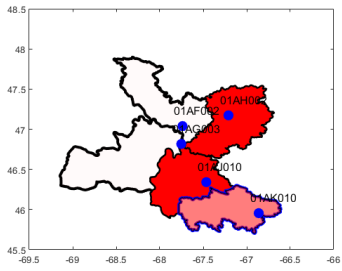
(a) *Scénario 12*



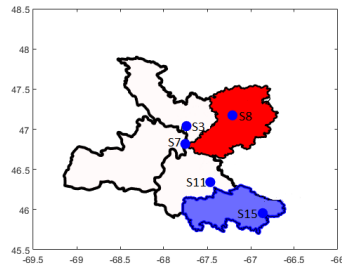
(b) *Scénario 14*



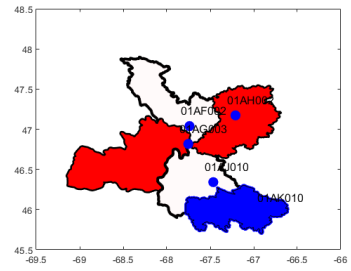
(c) *Scénario 7*



(d) *Scénario 10*



(e) *Scénario 4*



(f) *Scénario 9*

FIGURE B4 – *Scénarios possibles pour la station 15 obtenus par régression quantile (suite).*

C Code R

Dans les Tableaux C1 et C2, nous résumons les principales commandes du logiciel *R* utilisées pour générer les résultats de cette étude.

TABLE C1 – Commandes et Packages *R* utilisés pour générer les résultats (1).

Résultat	Commande {Package}	Exemple
Fig. 2.5	<i>BiCopKPlot</i> { <i>VineCopula</i> } (voir <i>Schepsmeier et al.</i> [2015])	<i>BiCopKPlot(pobs(resid.env1), pobs(resid.Prec1))</i> où <i>pobs</i> désigne la fonction décrite dans 1.2.1 pour obtenir la paire (U_{t_j}, V_{t_j}) .
Fig. 2.6, Fig. 3.6, Fig. A2 et Fig. A3	<i>contour</i> { <i>graphics</i> } (voir <i>Team et al.</i> [2017])	<i>contour(my_dist, pMvdc)</i> où <i>my_dist</i> est la distribution bivariable construite via la copule et les marges paramétriques et <i>pMvdc</i> la CDF correspondante.
Fig. 2.7	<i>rMvdc</i> { <i>copula</i> } (voir <i>Hofert et al.</i> [2014])	<i>rMvdc(n, my_dist)</i> , <i>n</i> étant le nombre d'observations générées.
Fig. 2.8	<i>boxplot</i> { <i>graphics</i> }	<i>boxplot(BOXPLOT[,1], BXPLOT[,2], BXPLOT[,7], BXPLOT[,10])</i> où chacun des éléments passés à la commande correspond à l'un des modèles de l'envasement proposés.
Fig. 3.5	<i>pairs.copuladata</i> { <i>VineCopula</i> }	<i>pairs.copuladata(U_volume)</i> , <i>U_volume</i> étant la matrice composée des distributions de la variable volume de chacune des stations.
Fig. 3.7	<i>chart.Correlation</i> { <i>PerformanceAnalytics</i> } (voir <i>Carl et al.</i> [2010])	<i>chart.Correlation(U_volume, method="kendall")</i> , le graphique peut représenter au choix l'un des coefficients de corrélation (Spearman, Pearson ou Kendall), une version avec histogramme peut être obtenue en ajoutant <i>histogram=TRUE</i> .

TABLE C2 – Commandes et Packages R utilisés pour générer les résultats (2).

Résultat	Commande {Package}	Exemple
Tab. A2	<i>summary</i> {base} (voir Kohl [2010])	<i>summary(APP1)</i> , APP1 un vecteur de données.
Tab. 2.2	<i>cor</i> {stats} (voir Team and Worldwide [2002])	<i>cor(vent,Env,method=c("kendall","spearman"))</i> .
Tab. 2.3	<i>auto.arima</i> {forecast} (voir Hyndman et al. [2020])	<i>auto.arima(resid.env, d = NA, D = NA, max.p = 5, max.q = 5, max.order = 5, max.d = 2, start.p = 0, start.q = 0, stationary = TRUE, seasonal = FALSE, ic = c("aicc", "aic", "bic"))</i> . La fonction renvoie le meilleur modèle ARIMA en fonction de la valeur AIC, AICc ou BIC. Elle effectue une recherche sur un modèle possible dans les limites des contraintes d'ordre fournies.
	<i>ks.test</i> et <i>Box.test</i> {stats}	<i>ks.test(resid.App1, "pnorm", mean = mean(resid.App1), sd = sd(resid.App1))</i> et <i>Box.test(resid.App1)</i> .
	<i>ad.test</i> {ADGofTest} (voir Bellosta and Bellosta [2009])	<i>ad.test(resid.App1, pnorm, mean = mean(resid.App1), sd = sd(resid.App1))</i> .
Tab. A4 et Tab. A5	<i>BiCopSelect</i> {VineCopula}	<i>BiCopSelect(pobs(resid.App1), pobs(resid.Tmoy1), familyset = c(3,4,5,104,204))</i> . Les copules choisies pour faire l'ajustement doivent être spécifiées au départ.
Tab. 2.4	<i>gofCopula</i> {copula}	<i>gofCopula(surJoeBiCopula(param=3.38), M1, N=50)</i> , M1 étant la matrice des paires de variables.
Tab. 3.3	<i>ad.test</i> {ADGofTest}	
Tab. 3.4	<i>fitdist</i> {fitdistrplus} (voir Delignette-Muller et al. [2020])	<i>fitdist(volume_15/mean(volume_15), "gamma")</i>
	<i>Box.test</i> {stats}	<i>Box.test(V_15, type="Ljung-Box")</i>
Tab. 3.5	<i>RVineStructureSelect</i> {VineCopula}	<i>RVineStructureSelect(U_volume, familyset = c(3,4,5), rotations = TRUE, indeptest = TRUE, level=0.05)</i> . Comme pour l'ajustement des copules aux paires de variables, le choix des copules qui seront ajustées aux différents arêtes de la structure R-vigne doivent être spécifiés au départ. Un test d'indépendance est mené au choix.
Tab. 3.6	<i>RVineVuongTest</i> {VineCopula}	<i>RVineVuongTest(U_volume, vine1, vine2)</i> . U_volume étant la matrice de données, vine1 et vine2 les deux structures ajustées à comparer.
Tab. 3.7	<i>h</i> et <i>hinverse</i> {vines} (voir Gonzalez-Fernandez et al. [2016])	<i>h(surGumbelCopula(1.65), j, l)</i> . j et l étant les quantiles spécifiques à chaque variable. La syntaxe est identique pour <i>hinverse</i> .

BIBLIOGRAPHIE

- Aas, K., and D. Berg, Models for construction of multivariate dependence—a comparison study, *The European Journal of Finance*, 15(7-8), 639–659, 2009.
- Aas, K., C. Czado, A. Frigessi, and H. Bakken, Pair-copula constructions of multiple dependence, *Insurance : Mathematics and economics*, 44(2), 182–198, 2009.
- Agence du bassin hydrologique de la Moulouya, <http://www.abhmoulouya.ma/>.
- Aghakouchak, A., Entropy–copula in hydrology and climatology, *Journal of Hydrometeorology*, 15(6), 2176–2189, 2014.
- Aguilera, R., R. Marcé, and S. Sabater, Detection and attribution of global change effects on river nutrient dynamics in a large Mediterranean basin., *Biogeosciences*, 12(13), 2015.
- Ait Hassou, L., F. Badaoui, O. G. Cyrille, A. Amar, A. Zoglat, and E. Ezzahid, Copulas for modeling the relationship between inflation and the exchange rate, in *International Work-Conference on Time Series Analysis*, pp. 217–228, Springer, 2017.
- Alahiane, N., A. El Mouden, A. A. Lhaj, and S. Boutaleb, Practical method proposed to estimate silting’s rate in small and hillside dams, *Journal of Water Resource and Protection*, 2014.
- Alina, B., Copula modeling for world’s biggest competitors, Ph.D. thesis, Master Thesis, Amsterdam School of Economics, Universiteit Van Amsterdam, 2018.
- Anderson, T. W., and D. A. Darling, A test of goodness of fit, *Journal of the American statistical association*, 49(268), 765–769, 1954.
- Arshad, M., M. Rasool, and M. Ahmad, Anderson Darling and modified Anderson Darling tests for generalized Pareto distribution, *Pakistan Journal of Applied Sciences*, 3(2), 85–88, 2003.
- Azur, M. J., E. A. Stuart, C. Frangakis, and P. J. Leaf, Multiple imputation by chained equations : What is it and how does it work ?, *International journal of methods in psychiatric research*, 20(1), 40–49, 2011.
- Bárdossy, A., and J. Li, Geostatistical interpolation using copulas, *Water Resources Research*, 44(7), 2008.
- Bárdossy, A., and G. Pegram, Copula based multisite model for daily precipitation simulation, *Hydrology & Earth System Sciences*, 13(12), 2009.

- Barlow, R., and F. Proschan, Statistical theory of reliability and life testing : Probability models, *Silver Spring, MD*, 1981.
- Bates, B., Z. Kundzewicz, and S. Wu, *Climate change and water*, Intergovernmental Panel on Climate Change Secretariat, 2008.
- Bedford, T., and R. M. Cooke, Probability density decomposition for conditionally dependent random variables modeled by vines, *Annals of Mathematics and Artificial intelligence*, 32(1-4), 245–268, 2001.
- Bedford, T., and R. M. Cooke, Vines : A new graphical model for dependent random variables, *Annals of Statistics*, pp. 1031–1068, 2002.
- Bellosta, C. J. G., and M. C. J. G. Bellosta, Package ADGofTest, 2009.
- Benaïcha, A. C., Étude des processus des transports solides et leur impact sur l'alluvionnement des barrages, *Mémoire de Magister, Université El Hadj Lakhdar–Batna*, 2011.
- Bernard, C., and C. Czado, Conditional quantiles and tail dependence, *Journal of Multivariate Analysis*, 138, 104–126, 2015.
- Bevacqua, E., D. Maraun, I. Hobæk Haff, M. Widmann, and M. Vrac, Multivariate statistical modelling of compound events via pair-copula constructions : analysis of floods in Ravenna (Italy), *Hydrology and Earth System Sciences*, 21(6), 2701–2723, 2017.
- Box, G. E., and G. M. Jenkins, Time series analysis : Forecasting and control Holden-Day, *San Francisco*, p. 498, 1970.
- Box, G. E., and D. A. Pierce, Distribution of residual autocorrelations in autoregressive-integrated moving average time series models, *Journal of the American statistical Association*, 65(332), 1509–1526, 1970.
- Brechmann, E., Truncated and simplified regular vines and their applications, *Munich, Germany : Center of Mathematics, Technical University of Munich*, 2010.
- Brechmann, E. C., K. Hendrich, and C. Czado, Conditional copula simulation for systemic risk stress testing, *Insurance : Mathematics and Economics*, 53(3), 722–732, 2013.
- Carl, P., B. G. Peterson, and M. B. G. Peterson, Package PerformanceAnalytics, Retrieved March, 29, 2011, 2010.
- Celeste, A. B., and M. Billib, Evaluation of stochastic reservoir operation optimization models, *Advances in Water Resources*, 32(9), 1429–1443, 2009.
- Chan, K. H., J. C. Hayya, and J. K. Ord, A note on trend removal methods : The case of polynomial regression versus variate differencing, *Econometrica*, Vol. 45, No. 3, pp. 737-744, 1977.
- Chang, B., and H. Joe, Prediction based on conditional distributions of vine copulas, *Computational Statistics & Data Analysis*, 139, 45–63, 2019.
- Chu, B., Recovering copulas from limited information and an application to asset allocation, *Journal of Banking & Finance*, 35(7), 1824–1842, 2011.
- Coles, S. G., and J. A. Tawn, Modelling extreme multivariate events, *Journal of the Royal Statistical Society : Series B (Methodological)*, 53(2), 377–392, 1991.
- Colosimo, I., P. L. de Vet, D. S. van Maren, A. J. Reniers, J. C. Winterwerp, and B. C. van Prooijen, The impact of wind on flow and sediment transport over intertidal flats, *Journal of Marine Science and Engi-*

- neering*, 8(11), 910, 2020.
- Cooke, R. M., H. Joe, and K. Aas, Vines arise, in *Dependence modeling : Vine copula handbook*, pp. 37–71, World Scientific, 2010.
- Cooke, R. M., H. Joe, and B. Chang, Vine regression, *Resources for the Future Discussion Paper*, pp. 15–52, 2015.
- Czado, C., Analyzing dependent data with vine copulas, *Lecture Notes in Statistics*, Springer, 2019.
- Czado, C., S. Jeske, and M. Hofmann, Selection strategies for regular vine copulae, *Journal de la Société Française de Statistique*, 154(1), 174–191, 2013.
- Dahan, S. R., and J. G. Grijzen, Managing urban water scarcity in Morocco : summary report (English), Washington, D.C. : World Bank Group, <http://documents.worldbank.org/curated/en/416241516117427311/summary-report>, 2017.
- De Michele, C., and G. Salvadori, A generalized Pareto intensity-duration model of storm rainfall exploiting 2-copulas, *Journal of Geophysical Research : Atmospheres*, 108(D2), 2003.
- Delignette-Muller, M. L., C. Dutang, R. Pouillot, J.-B. Denis, and M. M. L. Delignette-Muller, Package *fitdistrplus*, 2020.
- Diani, K., H. Tabyaoui, I. Kacimi, F. El Hammichi, and C. Nakhcha, Stream network modelling from aster gdem using archydro gis : application to the upper moulouya river basin (eastern, morocco), *Journal of Geoscience and Environment Protection*, 5(05), 1, 2017.
- Dißmann, J., Statistical inference for regular vines and application, 2010.
- Dißmann, J., E. C. Brechmann, C. Czado, and D. Kurowicka, Selecting and estimating regular vine copulae and application to financial returns, *Computational Statistics & Data Analysis*, 59, 52–69, 2013.
- Driouech, F., G. Mahé, M. Deque, C. Dieulin, T. El Heirech, and M. Milano, Evaluation d’impacts potentiels de changements climatiques sur l’hydrologie du bassin versant de la Moulouya au Maroc, *IAHS publication*, pp. 561–567, 2010.
- El Adlouni, S., Quantile regression C-vine copula model for spatial extremes, *Natural Hazards*, 94(1), 299–317, 2018.
- El-Jabi, N., and J. Rousselle, *Hydrologie fondamentale*, Éditions de l’École polytechnique de Montréal, 1987.
- Erhardt, T. M., C. Czado, and U. Schepsmeier, R-vine models for spatial time series with an application to daily mean temperature, *Biometrics*, 71(2), 323–332, 2015.
- Eschenburg, P., Properties of extreme-value copulas, Ph.D. thesis, Technische Universitat Munchen Fakultat fur Mathematik, 2013.
- Farrell, P. J., and K. Rogers-Stewart, Comprehensive study of tests for normality and symmetry : extending the Spiegelhalter test, *Journal of Statistical Computation and Simulation*, 76(9), 803–816, 2006.
- Fermanian, J.-D., Recent developments in copula models, *Econometrics*, 5(3), 34, 2017.
- Fink, H., Y. Klimova, C. Czado, and J. Stöber, Regime switching vine copula models for global equity and volatility indices, *Econometrics*, 5(1), 3, 2017.
- Geidosch, M., and M. Fischer, Application of vine copulas to credit portfolio risk modeling, *Journal of Risk*

- and *Financial Management*, 9(2), 4, 2016.
- Genest, C., and J.-C. Boies, Detecting dependence with Kendall plots, *The American Statistician*, 57(4), 275–284, 2003.
- Genest, C., and A.-C. Favre, Everything you always wanted to know about copula modeling but were afraid to ask, *Journal of hydrologic engineering*, 12(4), 347–368, 2007.
- Genest, C., B. Rémillard, and D. Beaudoin, Goodness-of-fit tests for copulas : A review and a power study, *Insurance : Mathematics and economics*, 44(2), 199–213, 2009.
- Golian, S., B. Saghafian, and A. Farokhnia, Copula-based interpretation of continuous rainfall–runoff simulations of a watershed in northern Iran, *Canadian Journal of Earth Sciences*, 49(5), 681–691, 2012.
- Gómez, M., M. C. Ausín, and M. C. Domínguez, Vine copula models for predicting water flow discharge at King George Island, Antarctica, *Stochastic Environmental Research and Risk Assessment*, 32(10), 2787–2807, 2018.
- Gonzalez-Fernandez, Y., M. Soto, J. Meys, and M. Y. Gonzalez-Fernandez, Package vines, *Vine*, 5, 1, 2016.
- Gyasi-Agyei, Y., Copula-based daily rainfall disaggregation model, *Water Resources Research*, 47(7), 2011.
- Hao, Z., and V. P. Singh, Entropy-copula method for single-site monthly streamflow simulation, *Water Resources Research*, 48(6), 2012.
- Hao, Z., and V. P. Singh, Modeling multisite streamflow dependence with maximum entropy copula, *Water Resources Research*, 49(10), 7139–7143, 2013.
- Hennemuth, B., S. Bender, K. Bülow, N. Dreier, E. Keup-Thiel, O. Krüger, C. Mudersbach, C. Radermacher, and R. Schoetter, Statistical methods for the analysis of simulated and observed climate data : applied in projects and institutions dealing with climate change impact and adaptation, in *CSC Report 13*, CSC Report 13, 2013.
- Hernandez, J. A., Vine copula modelling of dependence and portfolio optimization with application to mining and energy stock return series from the Australian market, 2015.
- Hofert, M., I. Kojadinovic, M. Maechler, J. Yan, M. M. Maechler, and M. Suggs, Package copula, <http://ie.archive.ubuntu.com/disk1/disk1/cran.r-project.org/web/packages/copula/copula.pdf>, 2014.
- Hosseinzadeh, A., H. SeyedKaboli, H. Zareie, A. Akhondali, and B. Farjad, Impact of climate change on the severity, duration, and frequency of drought in a semi-arid agricultural basin, *Geoenvironmental Disasters*, 2(1), 1–9, 2015.
- Hutchinson, T. P., Continuous bivariate distributions emphasising applications, *Tech. rep.*, 1990.
- Hyndman, R. J., G. Athanasopoulos, C. Bergmeir, G. Caceres, L. Chhay, M. O’Hara-Wild, F. Petropoulos, and S. Razbash, Package forecast, <https://cran.r-project.org/web/packages/forecast/forecast.pdf>, 2020.
- Ivanov, E., A. Min, and F. Ramsauer, Copula-based factor models for multivariate asset returns, *Econometrics*, 5(2), 20, 2017.
- Jackson, R. B., C. Stephen R, D. Clifford N, M. Diane M, N. Robert J, P. Sandra L, and R. Steven W, Water in a changing world, *Ecological applications*, 11(4), 1027–1045, 2001.

- Joe, H., Multivariate extreme-value distributions with applications to environmental data, *Canadian Journal of Statistics*, 22(1), 47–64, 1994.
- Joe, H., Families of m-variate distributions with given margins and m (m-1)/2 bivariate dependence parameters, *Lecture Notes-Monograph Series*, pp. 120–141, 1996.
- Joe, H., *Dependence modeling with copulas*, CRC press, 2014.
- Joe, H., and D. Kurowicka, *Dependence modeling : vine copula handbook*, World Scientific, 2011.
- Koenker, R., *Quantile Regression*, Cambridge University Press, 2005.
- Koenker, R., and G. Bassett Jr, Regression quantiles, *Econometrica : journal of the Econometric Society*, pp. 33–50, 1978.
- Kohl, M., The R Base package, 2010.
- Kraus, D., and C. Czado, D-vine copula based quantile regression, *Computational Statistics & Data Analysis*, 110, 1–18, 2017.
- Krouthén, J., Extreme joint dependencies with copulas A new approach for the structure of c-vines, 2015.
- Kwiatkowski, D., P. C. Phillips, P. Schmidt, and Y. Shin, Testing the null hypothesis of stationarity against the alternative of a unit root : How sure are we that economic time series have a unit root?, *Journal of econometrics*, 54(1-3), 159–178, 1992.
- Lagnoux, A., Renforcement statistique : Séries chronologiques, *ISMAG, Master1-MI0B246X, Université de Toulouse Le Mirail*, 53p, 2015a.
- Lagnoux, A., Séries chronologiques, *ISMAG, Master1-MI00141X, Université de Toulouse Le Mirail*, 53p, 2015b.
- Langat, P. K., L. Kumar, and R. Koech, Identification of the most suitable probability distribution models for maximum, minimum, and mean streamflow, *Water*, 11(4), 734, 2019.
- Lee, T., and J. D. Salas, Using copulas for stochastic streamflow generation, in *World Environmental and Water Resources Congress 2008 : Ahupua’A*, pp. 1–10, 2008.
- Les Landes, A. A., L. Aquilina, J. De Ridder, L. Longuevergne, C. Pagé, and P. Goderniaux, Investigating the respective impacts of groundwater exploitation and climate change on wetland extension over 150 years, *Journal of Hydrology*, 509, 367–378, 2014.
- Li, L., P. Liu, D. E. Rheinheimer, C. Deng, and Y. Zhou, Identifying explicit formulation of operating rules for multi-reservoir systems using genetic programming, *Water resources management*, 28(6), 1545–1565, 2014.
- Liu, P., S. Guo, X. Xu, and J. Chen, Derivation of aggregation-based joint operating rule curves for cascade hydropower reservoirs, *Water resources management*, 25(13), 3177–3200, 2011.
- Ljung, G. M., and G. E. Box, On a measure of lack of fit in time series models, *Biometrika*, 65(2), 297–303, 1978.
- Manner, H., Estimation and model selection of copulas with an application to exchange rates, *METEOR research memorandum 07/056, Maastricht University*, 2007.
- Manner, H., *Modeling asymmetric and time-varying dependence*, Datawyse/Universitaire Pers Maastricht, 2010.

- Mediero, L., A. Jiménez-Álvarez, and L. Garrote, Design flood hydrographs from the relationship between flood peak and volume, *Hydrology and Earth System Sciences*, 14(12), 2495, 2010.
- Monhor, D., A Chebychev inequality for multivariate normal distribution, *Probability in the Engineering and Informational Sciences*, 21, pp 289-300, 2007.
- Morris, G. L., and J. Fan, *Reservoir sedimentation handbook : Design and management of dams, reservoirs, and watersheds for sustainable use*, McGraw Hill Professional, 1998.
- Nápoles, O. M., *Bayesian belief nets and vines in aviation safety and other applications*, Citeseer, 2009.
- NASA, <https://power.larc.nasa.gov/data-access-viewer/>.
- Nash, J. E., and J. V. Sutcliffe, River flow forecasting through conceptual models part I- A discussion of principles, *Journal of hydrology*, 10(3), 282–290, 1970.
- Nelsen, R. B., Properties and applications of copulas : A brief survey, in *Proceedings of the first brazilian conference on statistical modeling in insurance and finance*, pp. 10–28, Citeseer, 2003.
- Nelsen, R. B., *An introduction to copulas*, Springer Science & Business Media, 2007.
- Newton, B., and B. C. Burrell, The April–May 2008 flood event in the Saint John River basin : Causes, assessment and damages, *Canadian Water Resources Journal/Revue canadienne des ressources hydriques*, 41(1-2), 118–128, 2016.
- Okou, G. C., Estimation du risque financier par l’approche de peaks over threshold (POT) et de la théorie des copules, Ph.D. thesis, Université Mohammed V-Agdal, Faculté des Sciences, Rabat, 2014.
- Oliveira, R., and D. P. Loucks, Operating rules for multireservoir systems, *Water resources research*, 33(4), 839–852, 1997.
- Orth, F., Die verlandung von staubecken, Ph.D. thesis, Technische Hochschule zu Berlin, 1934.
- Palaro, H. P., and L. K. Hotta, Using conditional copula to estimate value at risk, *Journal of Data Science*, 4, 93–115, 2006.
- Patton, A. J., Modelling time-varying exchange rate dependence using the conditional copula, *UCSD Discussion Paper*, 2001.
- Pereira, G., A. Veiga, T. Erhardt, and C. Czado, Spatial R-vine copula for streamflow scenario simulation, in *2016 Power Systems Computation Conference (PSCC)*, pp. 1–7, IEEE, 2016.
- Perreault, L., and B. Bobée, *Loi généralisée des valeurs extrêmes : Propriétés mathématiques et statistiques : Estimation des paramètres et des quantiles XT de période de retour T*, R350, INRS-Eau, 1992.
- Pickands, J., Multivariate extreme value distribution, *Proceedings 43th, Session of International Statistical Institution*, 1981.
- Pramanik, N., R. K. Panda, and D. Sen, Development of design flood hydrographs using probability density functions, *Hydrological Processes : An International Journal*, 24(4), 415–428, 2010.
- Pujol, N., L. Neppel, and R. Sabatier, Modélisation multivariée des précipitations max-annuelles par copules pour l’étude de stationnarité des pluies méditerranéennes françaises., *XIII World Water Congress*, 2008.
- Razali, N. M., and Y. B. Wah, Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests, *Journal of statistical modeling and analytics*, 2(1), 21–33, 2011.
- Remini, B., and W. Hallouche, Prévision de l’envasement dans les barrages du Maghreb, *LARHYSS Journal*

- P-ISSN 1112-3680/E-ISSN 2602-7828*, (4), 2005.
- Rosenblatt, M., Remarks on a multivariate transformation, *The annals of mathematical statistics*, 23(3), 470–472, 1952.
- Said, S. E., and D. A. Dickey, Testing for unit roots in autoregressive-moving average models of unknown order, *Biometrika*, 71(3), 599–607, 1984.
- Saidi, A., Erosion spécifique et prévision de l’envasement [Specific erosion and forecasting siltation], in *Colloque sur l’érosion des sols et l’envasement des barrages. Alger Décembre*, pp. 204–226, 1991.
- Samaniego, L., A. Bárdossy, and R. Kumar, Streamflow prediction in ungauged catchments using copula-based dissimilarity measures, *Water Resources Research*, 46(2), 2010.
- Schepsmeier, U., Efficient information based goodness-of-fit tests for vine copula models with fixed margins : A comprehensive review, *Journal of Multivariate Analysis*, 138, 34–52, 2015.
- Schepsmeier, U., et al., Package VineCopula, *R package version*, 2(5), 2015.
- Singh, V. P., *Entropy theory and its application in environmental and water engineering*, John Wiley & Sons, 2013.
- Singh, V. P., and L. Zhang, Copula–entropy theory for multivariate stochastic modeling in water engineering, *Geoscience Letters*, 5(1), 6, 2018.
- Singla, S., Impact du changement climatique global sur les régimes hydroclimatiques au Maroc : Tendances, ruptures et effets anthropiques sur les écoulements. Mémoire de Master : Recherche Eau et Environnement : Biol., géol., Agroressources Environnement, *Fonds IRD; Université de Montpellier 2*, https://horizon.documentation.ird.fr/exl-doc/pleins_textes/divers16-08/010055323.pdf, 2009.
- Sklar, A., Fonctions de répartition à n dimensions et leurs marges, *Publ. inst. statist. univ. Paris*, 8, 229–231, 1959.
- Snoussi, M., Review of certain basic elements for the assessment of environmental flows in the lower Moulouya, *IUCN International Union for Conservation of Nature, Gland*, 2004.
- Sugimoto, T., A. Bárdossy, G. G. Pegram, and J. Cullmann, Investigation of hydrological time series using copulas for detecting catchment characteristics and anthropogenic impacts, *Hydrology and Earth System Sciences*, 20(7), 2705, 2016.
- Tawn, J. A., Bivariate extreme value theory : Models and estimation, *Biometrika*, 75(3), 397–415, 1988.
- Team, R. C., and C. Worldwide, The R stats package, *R Foundation for Statistical Computing, Vienna, Austria : Available <http://www.R-project.org>*, 2002.
- Team, R. C., et al., The R Graphics package, 2017.
- Tekken, V., and J. P. Kropp, Climate-driven or human-induced : indicating severe water scarcity in the Moulouya river basin (Morocco), *Water*, 4(4), 959–982, 2012.
- Vanoni, V. A., Sedimentation engineering, in *American Society of Civil Engineers*, 2006.
- Vernieuwe, H., S. Vandenbergh, B. De Baets, and N. Verhoest, A continuous rainfall model based on vine copulas, *Hydrology and Earth System Sciences*, 19(6), 2685–2699, 2015.
- Vuong, Q. H., Likelihood ratio tests for model selection and non-nested hypotheses, *Econometrica : Journal*

of the Econometric Society, pp. 307–333, 1989.

- Xiong, L., K.-x. Yu, and L. Gottschalk, Estimation of the distribution of annual runoff from climatic variables using copulas, *Water Resources Research*, 50(9), 7134–7152, 2014.
- Yagbasan, O., Impacts of climate change on groundwater recharge in Küçük Menderes river basin in western Turkey, *Geodinamica Acta*, 28(3), 209–222, 2016.
- Yoo, C., and E. Cho, Vulnerability assessment of dam water supply capacity based on bivariate frequency analysis using copula, *Water*, 10(9), 1113, 2018.
- Young, G. K., Finding reservoir operating rules, *Journal of the Hydraulics Division*, 93(6), 297–322, 1967.
- Yu, J., X. Yao, W. Sun, and Z. Li, The impacts of climate change on water resources in the second Songhua river basin, China, in *IOP Conference Series : Earth and Environmental Science*, vol. 41, p. 012032, IOP Publishing, 2016.
- Zhang, J., Z. Li, X. Wang, X. Lei, P. Liu, M. Feng, S.-T. Khu, and H. Wang, A novel method for deriving reservoir operating rules based on flood classification-aggregation-decomposition, *Journal of Hydrology*, 568, 722–734, 2019.
- Zobeck, T., and R. Van Pelt, Erosion | Wind-induced, in *Encyclopedia of Soils in the Environment*, edited by D. Hillel, pp. 470–478, Elsevier, Oxford, doi :<https://doi.org/10.1016/B0-12-348530-4/00263-0>, 2005.