

ANALYSE DU COMPORTEMENT MULTIVARIE DE LA RESPONSE HYDRO-GEOMORPHIQUE VIA LES STATISTIQUES DE RANGS

Emna Gargouri-Ellouze¹, Rim Chérif^{1,2} et Julie Carreau³

¹ Université de Tunis El Manar, Ecole Nationale d'Ingénieurs de Tunis, LR99ES19 Laboratoire de Modélisation en Hydraulique et Environnement, 2092, Tunis, Tunisie

² : Université de Carthage, Institut Supérieur des Sciences et Techniques de l'Environnement

³ : HydroSciences Montpellier (UMR 5569), Université Montpellier 2 - Case MSE

Emna Gargouri-Ellouze, BP37, le Belvédère, 1002, Tunis, Tunisie. Emna.Gargouri@enit.rnu.tn;

Rim Chérif, BP37, le Belvédère, 1002, Tunis, Tunisie. rim.cherif@enit.rnu.tn

Julie Carreau, Place Eugène Bataillon - 34095 MONTPELLIER CEDEX 5. Julie.Carreau@univ-montp2.fr

Résumé :

L'analyse fréquentielle régionale est proposée pour estimer les quantiles de débits dans les bassins non jaugés. Pour régionaliser les bassins versants, les méthodes de partitionnement telles que l'analyse de cluster sont souvent appliquées. La délimitation des régions est basée sur des distances calculées entre les sites dans l'espace multidimensionnel : hydrologique, physiographique et géomorphologique.

Ce travail vise à construire des courbes de fréquence pluie-débit- régionales grâce aux copules. La méthode de classification hiérarchique est utilisée pour la délimitation des bassins hydrologiquement homogènes. Cette méthode utilise la distance modifiée de Mahalanobis, basée sur les rangs et le tau de Kendall.

Nous considérons dans ce travail de petits bassins versants situés dans la Dorsale tunisienne, suivis depuis 1992. Pour chaque région, le Kendall plot entre débit maximum et la pluie révèle que la dépendance existe et qu'elle est positive.

L'utilisation de la distance modifiée de Mahalanobis basée sur les rangs et le tau de Kendall, conduit à deux régions distinctes. Dans chaque région, nous avons construit une courbe de fréquence pluie-débit- régionale en utilisant la copule de Gumbel bivariée. La détermination des telles courbes régionales permet l'évaluation des caractéristiques de ruissellement qui représente un enjeu important dans la gestion des ressources en eau et aide à la prise de décision en matière de protection de l'environnement

Mots clés : analyse fréquentielle régionale, distance rangée modifiée de Mahalanobis, tau de Kendall, réponse hydro-géomorphologique.

1. Introduction

La communauté scientifique et la classe politique s'accordent à dire que les ressources et l'environnement de l'eau sont menacés (Bridging Divides for Water, 2009). Depuis déjà plusieurs années, les scientifiques affirment que les activités humaines ont perturbé les régimes naturels hydrologiques et écologiques (PHI, 2008). Ces perturbations ont affecté le climat, ce qui a provoqué des inondations et des sécheresses de plus en plus fréquentes et importantes dans plusieurs régions du globe que l'on croyait jusqu'ici protégées (GIEC, 2007). En outre, l'eau n'est plus garantie pour plusieurs peuples à travers le monde, l'écosystème est dégradé et la biodiversité est menacée.

Afin d'évaluer ces menaces, les hydrologues ont proposé des approches de décisions opérationnelles à différentes échelles spatio-temporelles et ont pu fournir des réponses adéquates aux acteurs socio-économiques et environnementaux (PHI-VI, 2002-2007). Cependant, ces approches ne sont performantes que si l'on dispose d'informations hydrométriques, à une échelle convenable, sur la région d'étude comme les bassins versants. Or, dans plusieurs régions du globe, les bassins versants sont peu ou pas du tout jaugés, et/ou dans certains cas, les réseaux de mesures existants fonctionnent mal, ou ne sont pas suivis, ou mal entretenus ou encore non étalonnés. La régionalisation est proposée pour répondre à ce manque d'information.

Elle est souvent utilisée en hydrologie, pour transférer des informations de sites jaugés vers d'autres non jaugés. Cette approche explore les relations et les interconnexions qui peuvent expliquer la réponse hydrologique, et permet ainsi de classer les bassins en régions avec un même comportement hydrologique que l'on appelle « voisins hydrologiques » (Ouarda et al., 2008). Il est à noter que ces régions ne sont pas nécessairement géographiquement contiguës. Ce transfert d'information est basé sur des mesures de similarité ou dissimilarité entre bassins. Celles-ci utilisent les attributs physiographiques et géomorphologiques des bassins et/ou les attributs hydro-climatiques. Ainsi, pour estimer les quantiles de débits dans les bassins non jaugés dans une région donnée, on fait appel à l'Analyse Fréquentielle Régionale (RFA).

Les deux principales étapes de la RFA sont :

- L'agrégation de bassins hydrologiquement homogènes en régions ;
- L'estimation régionale des quantiles.

La délimitation des bassins hydrologiquement homogènes en régions peut être réalisée par les méthodes de partitionnement telles que la classification hiérarchique. La délimitation des régions est basée sur des distances calculées entre les sites dans l'espace multidimensionnel : hydrologique, physiographique et géomorphologique (exemple : Burn et Goel, 2000, Weznah et al., 2013; Chérif et Bargaoui, 2013).

Dans la littérature plusieurs types de distances sont utilisées telles que : euclidienne, euclidienne quadratique, Manhattan, Chebyshev, Cosinus, Canberra, Minkowski et Mahalanobis.

Cunderlik et Burn (2006) recommande la distance de Mahalanobis pour évaluer la similarité entre les sites. En effet, cette distance considère la variance et la covariance des variables, ce qui n'est pas explicite avec les autres distances.

Dans ce travail nous adoptons la classification hiérarchique pour former les régions hydrologiquement homogènes en faisant appel à la distance de Mahalanobis modifiée, basée sur les rangs et le tau de Kendall. Dans chaque région homogène, les courbes de fréquence pluie-débit régionales sont établies grâce aux copules.

En effet, un engouement particulier ces dernières années a été constaté pour les copules car les distributions bivariées classiques ont l'inconvénient d'utiliser, des marginales identiques, ce qui n'est pas toujours vérifié pour les variables hydrologiques.

De plus, l'intérêt pour l'approche des copules consiste en leur possibilité de modéliser la structure de dépendance entre les variables indépendamment des lois marginales, la flexibilité dans le choix des marginales et l'extension à plusieurs variables (Sklar, 1959, Serinaldi et al., 2009).

Cet article est constitué de trois sections : après l'introduction, nous présentons dans la première section, les données et la méthodologie adoptée. La deuxième section est consacrée aux résultats et la dernière à la conclusion.

2. Données et méthodes

2.1. Données

Les bassins versants étudiés sont instrumentés depuis les années 1990 par la DGACTA¹. Ils sont situés entre la latitude 35° à 37° Nord et la longitude 8° à 11° Est. Ces bassins appartiennent à la dorsale. Ce sont de petits bassins versants de surfaces comprises entre 1km² et 10km², de relief relativement élevé. Leurs caractéristiques géomorphologiques (surface A , périmètre P , longueur du rectangle équivalent Lr , sa largeur lr , indice de pente globale Ip , indice Gravellus de compacité I_G , altitudes maximales $Alt Max$ et minimales $Alt min$ et dénivelée Ds) ont une variabilité importante (Tableau).

Tableau : Caractéristiques géomorphologiques des bassins (DGACTA, IRD)

Bassin	A (km ²)	P (km)	Alt Max (m)	Alt min (m)	Ds (m)	Ip (m/km)	Lr (km)	lr (km)	I_G
Abdeladhim	6.42	11.58	1224	1030	194	45	4.29	1.50	1.28
Abdessadok	3.07	7.95	1189	815	374	128	2.93	1.05	1.27
Arara	7.08	13.85	1352	910	442	78	5.68	1.25	1.46
BrahimZaher	4.64	12.80	1015	570	445	80	5.57	0.83	1.66
Dékikira	3.07	7.65	479	380	99	37	2.68	1.15	1.22
Echar	9.17	15.50	1190	970	220	35	6.29	1.46	1.43
El hanech	3.95	9.55	834	447	387	104	3.71	1.06	1.35
El Maleh	0.85	4.13	144	90	54	36	1.49	0.57	1.25
El mouidhi	2.66	7.70	363	235	128	43	2.95	0.90	1.32
Es séghir	4.31	10.05	231	70	161	41	3.93	1.10	1.36
Es sénéga	3.63	8.48	883	618	265	87	3.05	1.19	1.25
Fidh Ali	4.12	8.60	444	335	109	38	2.86	1.44	1.19
Fidhbenaceur	1.69	5.75	462	350	112	55	2.05	0.82	1.24
Hadada	4.69	9.90	1246	900	346	94	3.67	1.28	1.28
Janet	5.21	12.95	1191	820	371	67	5.53	0.94	1.59
Kamech	2.45	7.25	203	95	108	40	2.72	0.90	1.30
MrichetAnza	1.58	5.50	730	590	140	72	1.93	0.82	1.23
Mrira	6.13	12.30	940	770	170	35	4.9	1.25	1.39
Saadine	2.72	8.28	552	245	307	93	3.32	0.82	1.40
Saddine 1	3.84	9.70	1250	842	5	106	3.85	1.00	1.39
Saddine 2	6.53	16.80	1261	825	436	58	7.53	0.87	1.84
Sbahia	3.24	7.38	473	300	173	77	2.24	1.44	1.15
μ	4.14	9.71	802.55	554.86	229.36	65.86	3.78	1.07	1.36
Cv	0.48	0.34	0.50	0.56	0.60	0.42	0.41	0.24	0.12

Les données hydrométriques sont composées de séries de pluies (hauteur (mm), intensités (mm/h), durée (h)) et de crues (volume, débit de pointe, débit moyen, temps de pointe et temps de base).

Cette base de données hydro-géomorphologique constitue une matrice composée de vecteurs correspondants aux bassins versants. Chaque ligne de cette matrice représente un bassin avec ses différents attributs.

¹ DGACTA : Direction Générale des Aménagements et Conservation des Terres Agricoles

2.2. Méthodes

Nous adoptons dans ce travail la RFA pour déterminer les quantiles de débits. Cette approche se base sur :

- la délimitation des bassins hydrologiquement homogènes en régions grâce à la classification hiérarchique basée sur la distance de Mahalanobis modifiée, basée sur les rangs et le tau de Kendall
- l'estimation régionale des quantiles par les copules.

2.2.1. Distance de Mahalanobis

Nous proposons d'utiliser la distance de Mahalanobis modifiée, basée sur les rangs et le tau de Kendall. Cette méthode consiste à remplacer les variables d'étude par leurs rangs et la matrice de covariance par une matrice de corrélation.

La distance de Mahalanobis $d(X_i, X_j)$ est définie comme étant la mesure de dissimilarité entre deux réalisations X_i et X_j d'un vecteur aléatoire gaussien X avec une matrice de covariance Σ .

$$d(X_i, X_j) = \sqrt{(X_i - X_j)^T \Sigma^{-1} (X_i - X_j)}$$

X_i : $i^{\text{ème}}$ vecteur de variables

Σ : matrice de covariance.

Nous modifions la distance de Mahalanobis en remplaçant Σ par une matrice de corrélation P . Ceci conduit bien à une distance puisque la matrice de corrélation est définie positive tout comme la matrice de covariance. Cependant, utiliser la distance de Mahalanobis basée sur la corrélation revient à supposer que les variables suivent une loi Normale standard. La distance résultante met le même poids aux différences dans toutes les dimensions (puisque toutes les variables sont supposées être à la même échelle) et pondère les différences croisées en fonction de la matrice de corrélation.

Nous estimons les corrélations entre chaque paires de variables avec le τ de Kendall ce qui permet de représenter des structures de dépendance non-linéaires. Cependant, l'usage de la matrice de corrélation pour décrire la structure de dépendance n'élimine pas l'hypothèse d'une structure de dépendance gaussienne.

Ekström (2011) démontre qu'on peut utiliser n'importe quel type de distribution à condition de transformer les variables pour avoir une distribution Normale. Ceci peut être réalisé par la fonction de distribution empirique qui est basée sur les rangs des observations.

La fonction de distribution empirique appliquée aux observations donne les rangs et suit la loi de distribution Uniforme sur l'intervalle [0; 1]. La distance de Mahalanobis définie avec la matrice de corrélation P est appliquée aux rangs des observations plutôt que directement sur les observations.

2.2.2. Modélisation avec les copules

Avant de modéliser la corrélation existant entre la pluie et le débit, il convient de vérifier leur dépendance. Cette étape peut être réalisée soit par une mesure et/ou de façon graphique. Ensuite, on modélise la corrélation par les copules. Le choix définitif de la copule qui modélisera le mieux la corrélation, est effectué après ajustement

Mesure de dépendance

Les mesures d'association utilisées (Genest & Favre, 2007) pour la caractérisation de la dépendance sont : le coefficient de corrélation linéaire de Pearson ; le coefficient de corrélation des rangs de Spearman et le coefficient de corrélation des rangs de Kendall.

Le coefficient de corrélation de Pearson mesure la dépendance linéaire entre deux variables aléatoires (VA). Pour l'étude d'une dépendance non linéaire, les deux derniers coefficients sont plus adaptés puisqu'ils ne dépendent que du rang de chaque observation par opposition au coefficient de corrélation de Pearson qui utilise les réalisations des variables. Le tau de Kendall est utilisé dans ce

travail. Il peut être utilisé pour révéler une dépendance de deux caractéristiques qualitatives, et permet d'ordonner les éléments d'un échantillon en fonction de ces caractéristiques (Prokhorov, 2002).

Par ailleurs, on peut détecter la dépendance grâce à une méthode graphique basée sur les statistiques des rangs : Kendall-plots (K-plots) développée par Genest & Boies (2003). Les articles d'Abberger (2005) et Silvapulle & Zhang (2006) montrent que ces méthodes peuvent détecter non seulement la corrélation moyenne ou globale entre les variables mais aussi la dépendance asymptotique.

Copules

Si on considère un couple de variables aléatoires (VA), X et Y de fonctions de répartition marginales : $F_X(x) = P(X \leq x)$ et $F_Y(y) = P(Y \leq y)$ et de fonction de répartition conjointe :

$$F(x, y) = P(X \leq x \text{ et } Y \leq y)$$

Le concept des copules est de travailler sur le couple $(F_X(x), F_Y(y))$ et non pas sur le couple (X, Y) . On utilise donc le vecteur aléatoire $(U = F_X(x), V = F_Y(y))$ dont les lois marginales suivent une loi uniforme sur $[0, 1]$. On appelle "copule" du couple (X, Y) la fonction de répartition bivariée C ayant des lois marginales uniformes sur $[0, 1]$. C est unique si X et Y sont continues, et sachant qu'elle peut être étendue au cas multivarié. Elle est telle que, pour tout $(x, y) \in \mathbb{R}^2$ (théorème de Sklar, 1959):

$$F(x, y) = C(F_X(x), F_Y(y)) = C(U, V)$$

La copule associée à (X, Y) est alors, pour tout $(u, v) \in [0, 1]^2$, donnée par:

$$C(u, v) = F(F_X^{-1}(u), F_Y^{-1}(v)) = P(U \leq u, V \leq v)$$

La copule vérifie les propriétés suivantes :

$C(u, v)$ est une fonction croissante pour chacune de ses composantes u et v :

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

Pour tout $u_1 \leq u_2$ et $v_1 \leq v_2$; on a :

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

La densité f de F s'exprime alors par:

$$f(x, y) = c(F_X(x), F_Y(y)) \times f_X(x) \times f_Y(y)$$

où f_X et f_Y sont les densités des variables X et Y et $c(u, v)$ est la densité de la copule définie par:

$$c(u, v) = \frac{\partial^2 C(u, v)}{\partial u \partial v}$$

Les distributions conditionnelles $F_{Y|X}(y)$ et $F_{X|Y}(x)$ s'expriment en fonction des copules par:

$$F_{Y|X}(y) = C_1(F_X(x), F_Y(y))$$

où $C_1(u, v) = \frac{\partial C(u, v)}{\partial u}$ est la copule conditionnelle.

La copule empirique, pour un échantillon de couple (x, y) de taille n , est estimée par:

$$C_{\text{emp}}(u, v) = \frac{\text{nombre de couples } (x_i, y_i) \text{ tels que } F_X(x_i) \leq u \text{ et } F_Y(y_i) \leq v}{n}$$

Il existe plusieurs catégories de copules (Nelsen, 1999, Kallenberg, 2008) : les copules non paramétriques et celles paramétriques. Parmi les copules appartenant à la première catégorie, la copule empirique, la copule indépendance, les copules périodiques introduites par Alfonsi & Brigo (2005) et les copules de Fréchet (« upper bound » copule maximale et « lower bound » copule minimale). Dans la catégorie des copules paramétriques, on distingue plusieurs familles : les copules archimédiennes (Schweizer & Sklar, 1983), les copules méta-elliptiques notamment la copule gaussienne, la copule de Cauchy et la copule Student, les copules exponentielle et contamination (Kallenberg, 2008), la copule « de survie » (survival copula) (Venter, 2002), les copules des valeurs extrêmes à savoir Joe BB5 (Genest & Favre, 2007) et Galambos (Ghoudi *et al.*, 1998). Les deux premières familles restent les plus utilisées, particulièrement en hydrologie.

3. Résultats

3.1. Délimitation des bassins hydrologiquement homogènes en régions

Afin d'analyser le comportement multivarié de la réponse des bassins, nous combinons dans la matrice des prédicteurs les variables géomorphologiques et hydrométriques. Cette combinaison sert à se soustraire du test d'homogénéité des régions obtenues par la classification hiérarchique. L'algorithme de celle-ci utilise la distance de Mahalanobis modifiée, basée sur les rangs et le tau de Kendall. La classification conduit à deux régions homogènes distinctes : R1 et R2 (Figure 1) qui serviront à l'analyse régionale de la relation pluie-débit. Le choix du seuil pour le calcul du nombre de classes est justifié par la petite taille de l'échantillon. Nous pouvons aller jusqu'à trois régions si nous

appliquons la méthode de Mardia *et al.* (1979) : le nombre de régions $k = \sqrt{n/2}$ (n : nombre d'individus).

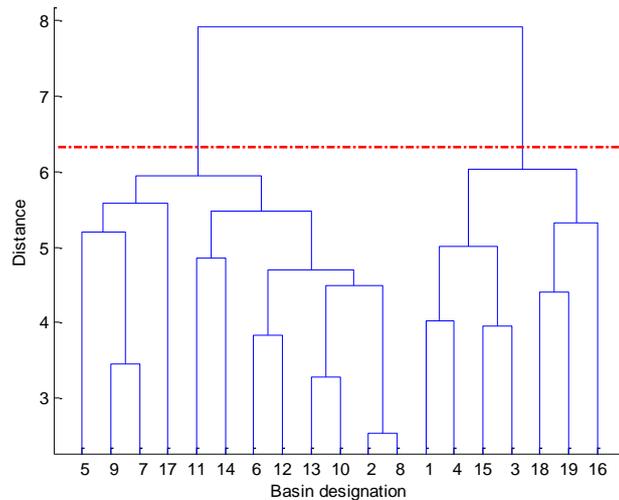


Figure 1: Dendrogramme des bassins

3.2. Modélisation par les copules

Pour chaque région, les K-plots entre le débit de pointe (Qmax) et l'intensité maximale (Imax) de la pluie montrent une dépendance positive avec des tau de Kendall de 0.322 et 0.365 dans R1 et R2 respectivement .

Le paramètre de la copule de Gumbel a été déterminé par le maximum de vraisemblance pour (Qmax, Imax) pour chacune des régions. Pour R1 les marginales de Qmax est la loi Log-Normale. Pour R2 Qmax suit la loi de Burr et Imax suit la loi Log-Normale. Par ailleurs, nous représentons les iso-lignes des couples simulés (Qmax, Imax) sur les Figure 2, Figure 3). Celles-ci correspondent aux courbes Qmax- Imax- Fréquence et qui donnent pour une fréquence fixée, les divers couples possibles de (Qmax, Imax).

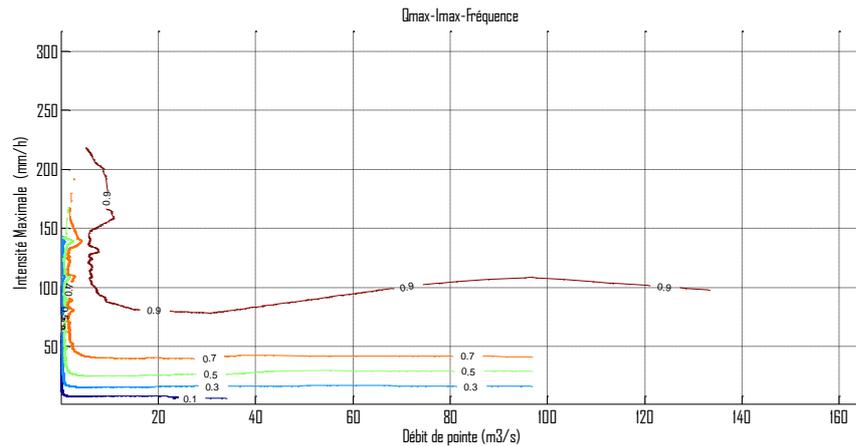


Figure 2: Qmax-Imax-Fréquence région 1

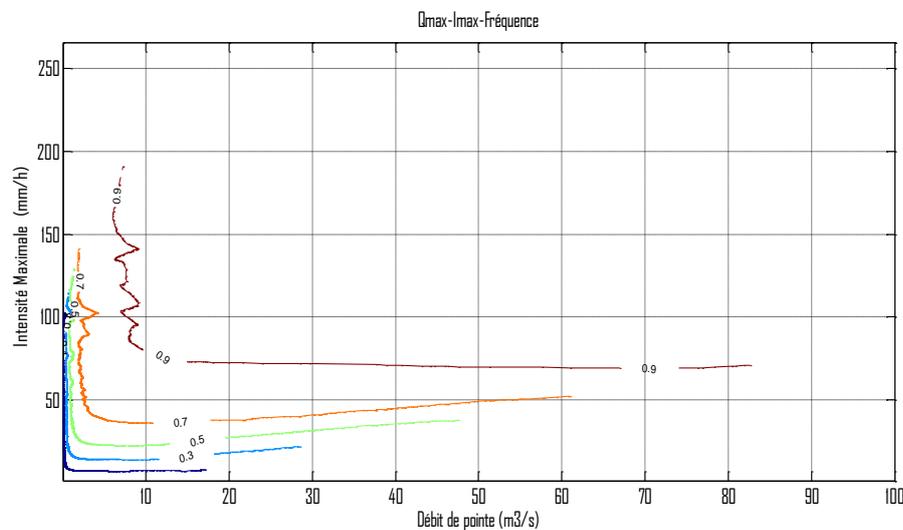


Figure 3 : Qmax-Imax-Fréquence région 2

4. Conclusion

La délimitation des bassins hydrologiquement homogènes en régions a été réalisée au moyen de la classification hiérarchique basée sur la distance de Mahalanobis modifiée. Dans cette dernière, les variables sont remplacées par leurs rangs et la matrice de covariance par une matrice de corrélation de Kendall. Deux régions ont été obtenues, à l'intérieur de chacune d'elles nous avons estimées les quantiles bivariés (Q_{max} , I_{max}) par la copule de Gumbel.

La détermination de Q_{max} connaissant I_{max} représente un important enjeu dans la gestion des ressources en eau et aide à la décision dans la gestion des ouvrages hydrauliques en cas de crue.

Bibliographie

- Abberger, K. (2005) A simple graphical method to explore tail-dependence in stock-return pairs. *Applied Financial Economics*, **15**, 43–51.
- Alfonsi, A. & D. Brigo (2005) New families of Copulas based on periodic functions. *Communications in Statistics: Theory and Methods*, **34** (7), 1437-1447.
- Bridging Divides for Water (2009) 5th World Water Forum, Istanbul, <http://www.worldwaterforum5.org/>.
- Burn, D. H. & Goel, N. K. (2000) The formation of groups for regional flood frequency analysis. *Hydrol. Sci. J.* **45** (1), 97-112

- Chérif R. and Bargaoui Z. (2013), Regionalization of maximum annual runoff using hierarchical and trellis methods and physiographic information. *Water Resources Management*: 27:2947-2963 DOI 10.1007/s11269-013-0325-0.
- Cunderlik, J.M. and D.H. Burn (2006). "Site-focussed nonparametric test of regional homogeneity based on flood regime", *Journal of Hydrology*, 318, 301-315
- Ekström, J. (2011). Mahalanobis' Distance Beyond Normal Distributions. *UCLA Stat, Preprints* 624.
- Genest, C. & Boies, J. C. (2003) Detecting dependence with Kendall plots. *Amer. Statist.* **57**, 275–284.
- Genest, C., Favre, A.C., (2007), Everything you always wanted to know about copula modeling but were afraid to ask. *J. Hydrol. Eng. ASCE* 12, 347–368.
- Ghoudi, K., A., Khoudraji & L.P., Rivest (1998) Propriétés statistiques des copules de valeurs extrêmes bidimensionnelles. *Rev Can Stat.* **26** (1), 187-197.
- GIEC (2007) Bilan 2007 des changements climatiques. Contribution des Groupes de travail I, II & III au quatrième Rapport d'évaluation du Groupe d'experts intergouvernemental sur l'évolution du climat [Équipe de rédaction principale, Pachauri, R.K. & Reisinger, A. (publié sous la direction de~)]. GIEC, Genève, Suisse, 103 pages.
- Kallenberg, W. C.M. (2008) Modelling dependence. *Insurance: Mathematics and Economics* **42**, 127–146.
- Kanti Mardia et al. (1979). *Multivariate Analysis*. Academic Press.
- Nelsen, R. B. (1999) *An Introduction to Copulas*. Springer, New York, USA.
- Ouarda, T.B.M.J., K.M. Bâ, C. Diaz-Delgado, A. Cârsteanu, K. Chokmani, H. Gingras, E. Quentin, E. Trujillo & B. Bobée (2008) Intercomparison of regional flood frequency estimation methods at ungauged sites for a Mexican case study. *J. Hydrol.* 348, 40- 58.
- Programme Hydrologique International (2008) Rapport Final, SC-2008/WS/31, UNESCO, Paris.
- Programme Hydrologique International, PHI-VI (2002-2007) Programme intergouvernemental de l'UNESCO sur les ressources en eau, http://www.unesco.org/water/ihp/ihp_six_fr.shtml.
- Prokhorov, A.V. (2002), Kendall coefficient of rank correlation. In: *Encyclopaedia of Mathematics.*, edited by Michiel Hazewinkel, CWI, Amsterdam, Netherlands.
- Schweizer, B. & Sklar, A. (1983) *Probabilistic Metric Spaces*. Elsevier, Amsterdam, The Netherlands.
- Serinaldi, F., B. Bonaccorso, A. Cancelliere & S. Grimaldi (2009) Probabilistic characterization of drought properties through copulas. *Physics and Chemistry of the Earth*. DOI 10.1016/j.pce.2008.09.004
- Silvapulle, P. & Zhang, X. (2006) *Assessing dependence changes in the Asian financial market returns using plots based on nonparametric measures*. Series Monash Econ. Busin. Statist. Working Papers 9-06, Monash University, Department of Econometrics and Business Statistics, Monash, Australia.
- Sklar, A. (1959) Fonctions de répartition à n dimensions et leurs marges. *Publ. Inst. Statist.* **8**, 229–231. Université Paris VIII, France.
- Venter, G. (2002) Tails of copulas. *Proc. Casualty Actuarial Society.* **89**, 68–113.
- Wazneh, H., F. Chebana and T. B. M. J. Ouarda (2013) Depth-based regional index-flood model. *Water Resources Research*, 49 (12), pages 7957–7972, DOI: 10.1002/2013WR013523