



UNIVERSITÉ DE MONTPELLIER

VALEURS EXTRÊMES

Statistique des événements extrêmes et applications

Thamara RENOIR

Lilou ZULEWSKI



STATISTIQUE
SCIENCE DES DONNÉES BIOSTATS
UNIVERSITÉ DE MONTPELLIER



27 Février 2024

Table des matières

Introduction	3
1 Analyse Univariée	4
1.1 Extraction des maxima	4
1.2 Ajustement d'une GEV sur les maxima	5
1.3 Niveaux de retour	8
2 Analyses bivariées	8
2.1 Calcul du quantile extrême conditionnel	9
2.2 Analyse des stations SA et 12 (SB)	10
2.3 Analyse des stations SA et 1 (SC)	13
2.4 Analyse des stations SA et 4 (SD)	14
2.5 Conclusion	16
Annexe	17

Résumé

En océanographie, l'étude des vagues extrêmes est d'une importance cruciale. En effet, celle-ci permet une meilleure gestion de la sécurité maritime et de la prévention des catastrophes afin d'évaluer les risques pour les infrastructures côtières. En s'appuyant sur la théorie des valeurs extrêmes, le présent travail modélise le comportement extrême des vagues dans le golfe du Lion, situé en Catalogne, et s'articule autour de deux approches : une première axée sur l'analyse univariée en se concentrant sur une station, suivie d'une analyse bivariée impliquant des paires de stations. Cette étude permettra une compréhension approfondie du comportement des vagues extrêmes sur ce site, donnant ainsi des informations essentielles pour la gestion des zones côtières.

Introduction

Dans ce travail, nous nous appuyons sur les données de vagues extrêmes du golfe du Lion contenues dans les fichiers `DonneesVagues.RData` et `DonneesStations.RData`. Les coordonnées géographiques des vingt stations d'étude sont indiquées dans la base de données `buoysInfos` : nous disposons de la longitude (*lon*), la latitude (*lat*), la profondeur (*depth*) et le nom de la station (*stationName*). Les hauteurs significatives de vagues horaires mesurées en chacune de ces stations sont présentes dans la base de données `DonneesVagues` : les mesures ont été effectuées quotidiennement et à toute heure de 1961 à 2012.

Voici un extrait de ces bases de données :

indexStation	lon	lat	depth	stationName
1	4.640	42.060	2326.11743	Lion
3	4.140	43.460	8.91505	EspigobsLa
4	6.707	43.296	146.92639	StMaximeLa
6	6.740	43.300	221.32100	StMaximeGr
7	7.800	43.400	2324.22217	61001

TABLE 1 – Extrait des coordonnées des stations d'étude

date	station1	station2	station3	station4	station5
1961-01-01 02 :00 :00	0.5120	0.1385	0.1775	0.1967	0.3012
1961-01-01 03 :00 :00	0.5465	0.0893	0.1448	0.1693	0.3608
1961-01-01 04 :00 :00	0.5895	0.0648	0.1429	0.1746	0.4378
1961-01-01 05 :00 :00	0.6452	0.0531	0.1728	0.2094	0.5271
1961-01-01 06 :00 :00	0.7143	0.0477	0.2211	0.2613	0.6136

TABLE 2 – Extrait des hauteurs de vagues dans les stations

L'analyse de ces hauteurs de vagues se fera en plusieurs parties. D'abord, nous allons effectuer une analyse univariée dans laquelle nous choisirons une station de référence pour calculer les quantiles extrêmes et discuter les niveaux de retour associés aux périodes de retour de 100, 500 et 1000 ans. Ensuite, nous répondrons à un objectif de prédiction des quantiles extrêmes conditionnels dans le cadre d'une analyse bivariée, notamment en considérant notre station de référence et une autre station. Nous comparerons les résultats obtenus dans différentes analyses bivariées. Enfin,

1 Analyse Univariée

Dans cette première partie, nous nous intéressons à l'analyse univariée des hauteurs de vagues extrêmes du golfe du Lion. Pour ce faire, nous considérons la station 6 comme station de référence, notée *SA*, sur laquelle nous allons mettre en place une approche par GEV (*Valeurs Extrêmes Généralisées*).

1.1 Extraction des maxima

La première étape de cette analyse univariée est l'extraction par blocs des maxima de la station *SA*. Nous considérons alors les maxima annuels de vagues que nous avons recueillis dans le vecteur suivant :

$$maxima_{\text{annuels}} = \begin{pmatrix} 7.012389 \\ 8.181569 \\ 6.392475 \\ \dots \\ 8.214052 \end{pmatrix}$$

Ce vecteur de maxima annuels est de taille $n = 52$. Ce nombre de maxima annuels est simplement égal au nombre d'années disponibles dans les données utilisées.

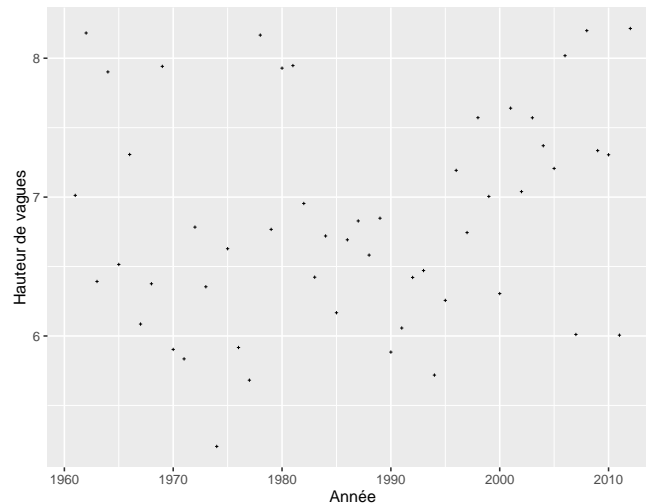


FIGURE 1 – Maxima annuels de hauteur de vagues de la station 6 (SA) du golfe du Lion

Sur la figure 1, il n'est pas évident de confirmer ou infirmer l'hypothèse de stationnarité. Ainsi, nous allons ajuster une GEV dans chacun des cas pour vérifier la réponse à ce postulat.

1.2 Ajustement d'une GEV sur les maxima

Stationnarité des données.

D'abord, nous supposons que nos données sont stationnaires voire indépendantes. Dès lors, nous pouvons ajuster une GEV sur les maxima annuels de la station de référence *SA* à l'aide de la fonction `gev.fit()` de R. Nous obtenons les estimateurs présentés dans le tableau 3 suivant.

Paramètre	Estimation	Ecart-type	Intervalle de confiance à 95%
μ	6.560	0.119	[6.326; 6.794]
σ	0.735	0.089	[0.560; 0.910]
γ	-0.258	0.139	[-0.530; 0.015]

TABLE 3 – Estimateurs des paramètres du modèle

Dès lors, nous supposons que les données ne sont pas stationnaires et nous ajoutons une tendance linéaire en fonction du temps au paramètre μ du modèle. Nous supposons ainsi qu'il existe β_0 et β_1 tels que $\mu(t) = \beta_0 + \beta_1 t$. Dans ce cas, les estimations des paramètres obtenus en utilisant la fonction `gev.fit()` sont indiqués dans le tableau 4.

Paramètre	Estimation	Ecart-type	Intervalle de confiance à 95%
β_0	6.300	0.231	[6.848; 6.753]
β_1	9.39×10^{-3}	7.147×10^{-3}	$[-4.612 \times 10^{-3}; 2.340 \times 10^{-2}]$
σ	0.707	8.342×10^{-2}	[0.543; 0.870]
γ	-0.213	1.297×10^{-1}	[-0.467; 0.041]

TABLE 4 – Estimateurs des paramètres du modèle avec tendance linéaire

Désormais, nous pouvons comparer ces deux modèles obtenus à l'aide d'un test de rapport de vraisemblance. L'hypothèse nulle de ce test est que le modèle restreint, à savoir le modèle construit avec hypothèse de stationnarité, est aussi bon que le modèle complet, supposant une linéarité, pour décrire les données. En d'autres termes, l'hypothèse nulle stipule qu'il n'y a pas de différence significative entre les deux modèles. Ici, la statistique de test obtenue est égale à $D = 1.736$ ce qui est inférieur au quantile d'ordre 0.95 d'une χ_1^2 : nous ne rejetons pas l'hypothèse nulle. Par conséquent, nous ne rejetons pas l'hypothèse qu'il n'y a pas de tendance linéaire sur μ .

Par la suite, nous considérerons alors que nos données sont stationnaires et nous retenons le premier modèle.

Ajustement d'un modèle Gumbel.

L'estimateur du paramètre γ étant négatif, nous en déduisons que la distribution des données de la station de référence est dans le domaine d'attraction d'une loi de Weibull. Cependant, 0 est dans l'intervalle de confiance de cet estimateur ce qui implique qu'il est envisageable d'ajuster une loi de Gumbel sur ces données.

En ajustant une telle loi sur la station de référence, nous obtenons les estimations des paramètres du tableau 5 suivantes :

Paramètre	Estimation	Ecart-type	Intervalle de confiance à 95%
μ	6.465	0.101	[6.458, 6.471]
σ	0.687	0.072	[0.682, 0.692]

TABLE 5 – Estimateurs des paramètres du modèle ajusté sur une loi de Gumbel

Nous pouvons comparer ce modèle de Gumbel estimé avec le modèle de Weibull construit ci-avant en appliquant à nouveau un test de rapport de vraisemblance. Ici, l'hypothèse nulle est que $\gamma = 0$. La log-vraisemblance des modèles de Gumbel et Weibull sont environ égales à -60.745 et -59.022 respectivement. De plus, la statistique de test D obtenue vaut 3.446 ce qui est inférieur au quantile $c_{0.05} = 3.841$ d'ordre 0.95 d'une χ^2_1 . Par conséquent, nous ne pouvons pas rejeter l'hypothèse nulle.

Nous pouvons comparer ces deux modèles en les confrontant graphiquement aux données de la station 6.

L'étude des différentes représentations graphiques de la figure 2 indique que le modèle de Weibull fournit un ajustement correct de nos données. De plus, les niveaux de retours estimés semblent en adéquation avec l'hypothèse $\gamma < 0$.

Sur la figure 3, nous constatons une légère amélioration du *Probability Plot* pour le modèle de Gumbel. À l'inverse, le *Quantile Plot* et le *Return Level Plot* sont moins bons que sous l'hypothèse d'une loi de Weibull. En effet, les points s'alignent moins correctement sur les droites des données de la station 6.

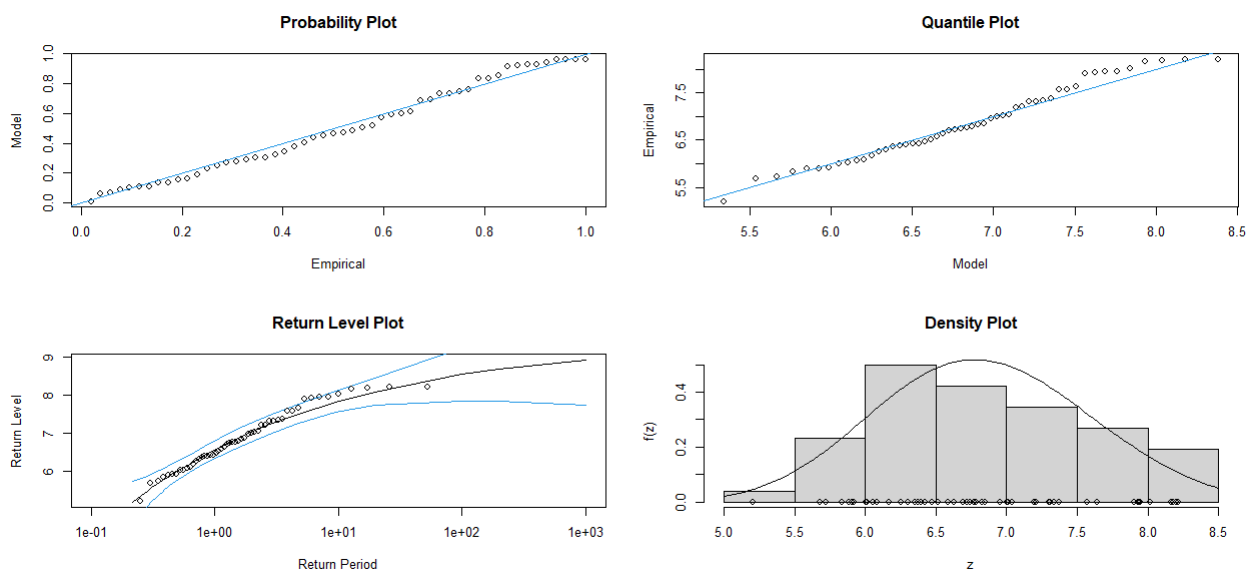


FIGURE 2 – Graphiques confrontant le modèle de Weibull estimé aux données de la station
SA

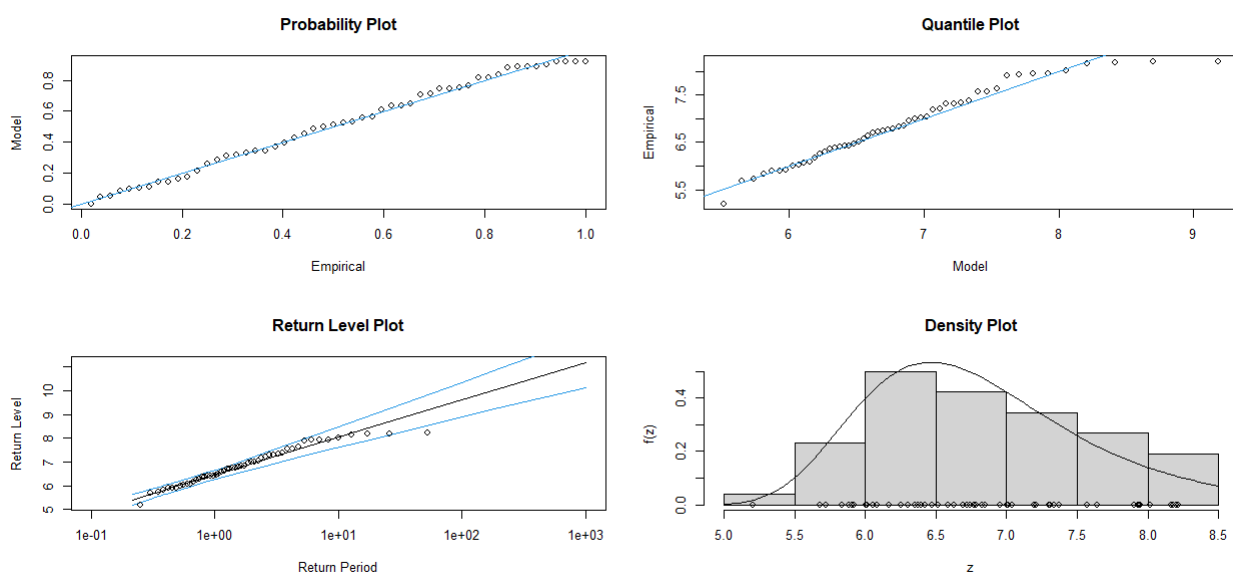


FIGURE 3 – Graphiques confrontant le modèle de Gumbel estimé aux données de la station
SA

D'après ces analyses numériques et graphiques, nous décidons de ne pas conserver le modèle Gumbel mais plutôt le modèle ajusté sur une loi de Weibull.

1.3 Niveaux de retour

Maintenant qu'un modèle est ajusté sur les données, nous pouvons calculer les niveaux de retour associés aux périodes de retour 100, 500 et 1000 ans. On rappelle que le niveau de retour $x_{\frac{1}{T}}$ associé à une période de retour T correspond à la hauteur maximale des vagues que nous nous attendons à dépasser en moyenne une fois toutes les T périodes.

Période de retour	Niveau de retour	Intervalle de confiance à 95%
100 ans	8.540	[7.845; 9.236]
500 ans	8.837	[7.778; 9.895]
1000 ans	8.931	[7.725; 10.137]

TABLE 6 – Niveaux de retour associés aux périodes de retour 100, 500 et 1000 ans

D'après la définition des niveaux de retour, il est cohérent que les niveaux de retour du tableau 6 augmentent avec la taille de la période. Cependant, nous remarquons que les niveaux de retour estimés sont supérieurs à ceux observés graphiquement sur le *Return Level Plot* ; de la figure 2.

Pour la station de référence, la hauteur maximale de vagues mesurée de 1961 à 2012, correspondant à 52 ans, est de 8.214 mètres arrondi à 10^{-3} : le quantile associé à la période de retour 100 n'a donc pas été atteint jusqu'en 2012 même si $8.214 \in [7.845; 9.236]$. Ici, cette constatation est consistante puisque la période de retour de 100 ans est largement supérieure à 50 ans. En revanche, nous pouvons nous attendre à observer des vagues de plus de 8.540 mètres de haut dans la station de référence d'ici les années à venir. D'après les niveaux de retour calculés, il est possible de voir des vagues extrêmes de 8.931 mètres de hautes d'ici les 1000 prochaines années.

2 Analyses bivariées

Dans cette seconde partie, nous nous intéressons à une analyse bivariée des hauteurs de vagues extrêmes du golfe de Lion : nous effectuerons une analyse bivariée entre les stations *SA* et 12 puis entre les stations *SA* et 1 et enfin, entre les stations *SA* et 4. Ici, l'objectif principal est de prédire les quantiles extrêmes conditionnels z_p vérifiant $P(X > z_p | Y > y) = p$ où p doit être petit : Y représente la hauteur de vagues de *SA* et X la hauteur de vagues à une autre station (4 ou 1 ou 12).

2.1 Calcul du quantile extrême conditionnel

Commençons par exprimer le quantile extrême conditionnel z_p en fonction de p .

Notons $G_{X,Y}$ la loi limite du vecteur $\left(\max_{1 \leq i \leq n} X_i, \max_{1 \leq i \leq n} Y_i\right)$ de marginales G_X et G_Y après normalisation, ainsi que F_X la fonction de répartition de X , F_Y celle de Y et $F_{X,Y}$ celle du couple (X, Y) .

Nous avons alors :

$$\begin{aligned} P(X > z_p | Y > y) &= \frac{P(X > z_p, Y > y)}{P(Y > y)} \\ &= \frac{1 - (F_X(z_p) + F_Y(y) - F_{X,Y}(z_p, y))}{1 - F_Y(y)} \\ &\approx \frac{1 - (G_X^{1/n}(z_p) + G_Y^{1/n}(y) - G_{X,Y}^{1/n}(z_p, y))}{1 - G_Y^{1/n}(y)} \end{aligned}$$

Dès lors, nous pouvons envisager deux cas de figure, celui de l'indépendance asymptotique et celui de l'égalité des marginales de $G_{X,Y}$.

Cas d'indépendance asymptotique.

Dans le cas d'indépendance asymptotique, nous pouvons écrire que $G_{X,Y} = G_X G_Y$ et

$$p = P(X > z_p | Y > y) \approx \frac{1 - [G_X^{1/n}(z_p)(1 - G_Y^{1/n}(y)) + G_Y^{1/n}(y)]}{1 - G_Y^{1/n}(y)}$$

Nous avons alors $z_p \approx G_X^{-1}((1 - p)^n)$.

Cas d'égalité des marginales.

A contrario, dans un cas d'égalité des marginales, nous avons $G = G_X = G_Y$ et nous pouvons alors écrire que $G_{X,Y} = G(\min(z_p, y))$.

Supposons que $z_p > y$, alors $p \approx \frac{1 - G^{1/n}(z_p)}{1 - G^{1/n}(y)}$ d'où $z_p \approx G^{-1}([1 - (1 - G^{1/n}(y))p]^n)$.

Remarque. Si $z_p \leq y$ alors $p = 1$, et il suffirait de prendre n'importe quel $z_p \leq y$. Cependant, nous prendrons dans la suite p petit sans nous intéresser à ce cas.

Afin de calculer les quantiles extrêmes conditionnels, il nous reste à savoir dans quel cas nous nous situons pour les couples de stations considérées (6, 12), (6, 1) et (6, 4).

2.2 Analyse des stations SA et 12 (SB)

Afin d'effectuer l'analyse bivariee des stations SA et 12 (notée SB), nous commençons par analyser cette dernière.

Analyse univariée de la station SB .

Cette station SB est distante d'environ 54 kilomètres de la station SA . Nous ajustons une GEV aux maxima annuels de la station SB pour laquelle les estimateurs des paramètres du modèle sont présents dans le tableau 7 suivant.

Paramètre	Estimation	Ecart-type	Intervalle de confiance à 95%
μ	3.114	0.072	[2.973; 3.256]
σ	0.463	0.052	[0.360; 0.567]
γ	0.034	0.098	[−0.157; 0.227]

TABLE 7 – Estimateurs des paramètres du modèle de la station SB

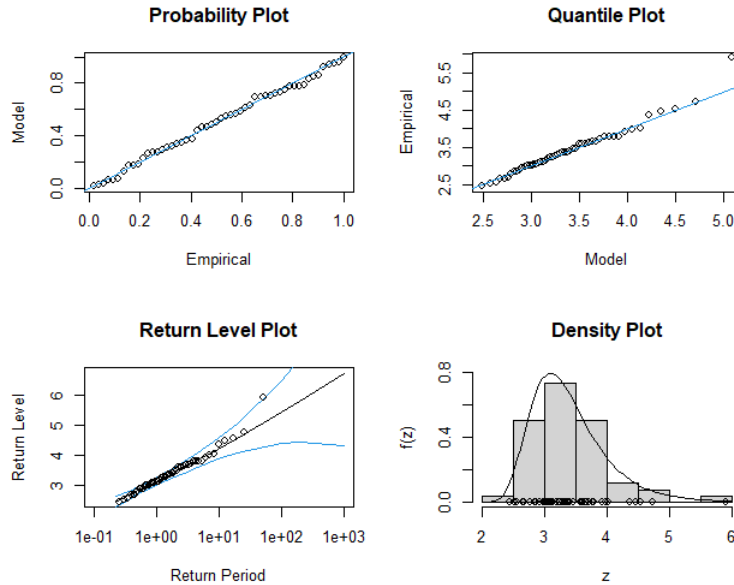


FIGURE 4 – Graphiques confrontant le modèle de Fréchet estimé aux données de la station SB

Nous remarquons dans le tableau 7 que l'estimateur de γ est positif : nous en déduisons alors que la distribution des données est dans le domaine d'attraction d'une loi de Fréchet. De plus, comme 0 est dans l'intervalle de confiance de γ , nous envisageons l'ajustement des données par une loi de Gumbel. Cependant, après avoir tester la nullité du paramètre γ par un test

du rapport de vraisemblance, nous rejetons finalement l'hypothèse $\gamma = 0$ et par conséquent, l'ajustement par une loi de Gumbel.

L'étude des graphiques de la figure 4 confirme que le modèle ajusté est correct.

Étude de la dépendance des marginales des stations SA et SB .

Maintenant, nous pouvons débiter l'analyse bivariable des stations SA et SB par l'étude de la dépendance entre les marginales de $G_{X,Y}$ où Y représente la hauteur des vagues de la station SB .

À première vue, la figure 5 nous indique que les maxima des deux stations ne sont pas dépendants. Nous allons étudier cette dépendance plus précisément à travers le couple $(\chi, \bar{\chi})$, qui caractérise la dépendance, et nous confirmons nos résultats en estimant l'indice extrême θ à l'aide de la fonction de dépendance de Pickands A .

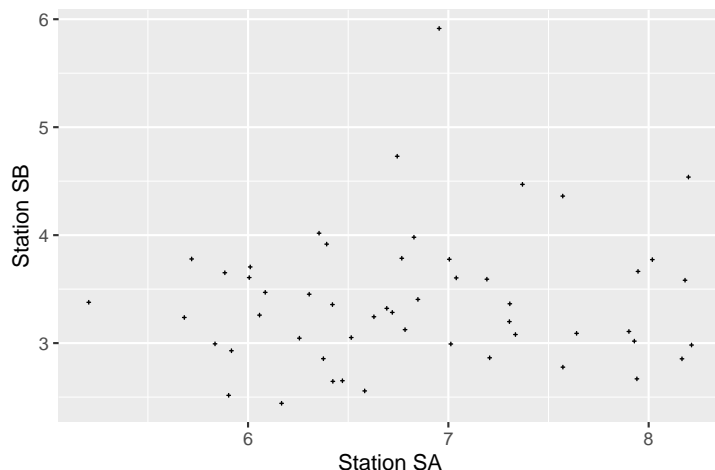


FIGURE 5 – Hauteurs maximales des vagues mesurées aux stations SA et SB

Nous traçons les fonctions $\chi(u)$ et $\bar{\chi}(u)$ calculées à partir de nos données en figure 6. Nous constatons que $\chi = \lim_{u \rightarrow 1} \chi(u) = 0$ et que $\bar{\chi} = \lim_{u \rightarrow 1} \bar{\chi}(u) < 1$: nous sommes donc dans un cas d'indépendance asymptotique. Afin de confirmer ce résultat, nous estimons l'indice extrême à l'aide de la fonction de dépendance de Pickands de la figure 7. Nous constatons sur cette figure que $\hat{A}(\frac{1}{2})$ est égal à 1 impliquant que $\hat{\theta} = 2$: nous sommes bien dans un cas d'indépendance asymptotique.

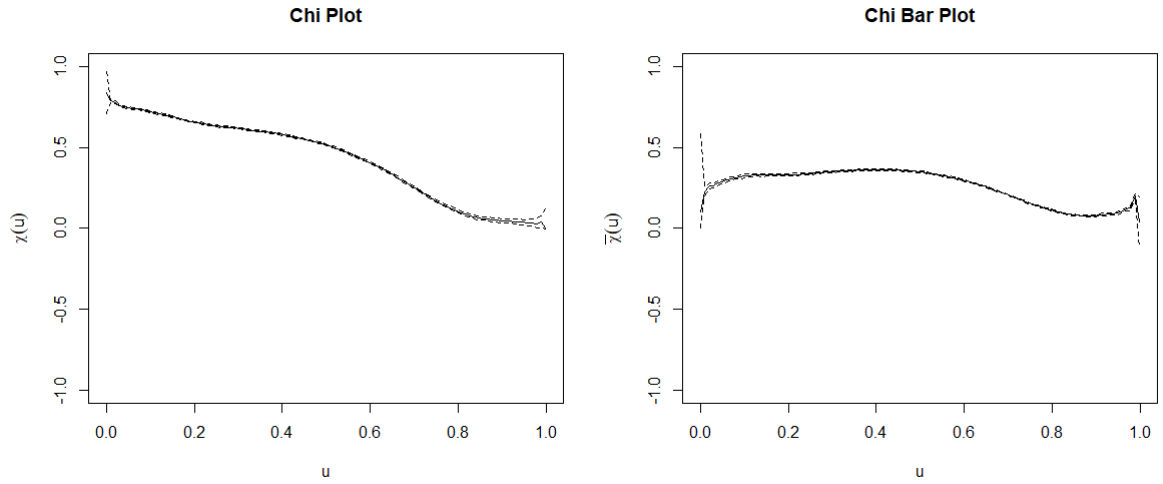


FIGURE 6 – Représentation graphique des fonctions $\chi(u)$ et $\bar{\chi}(u)$

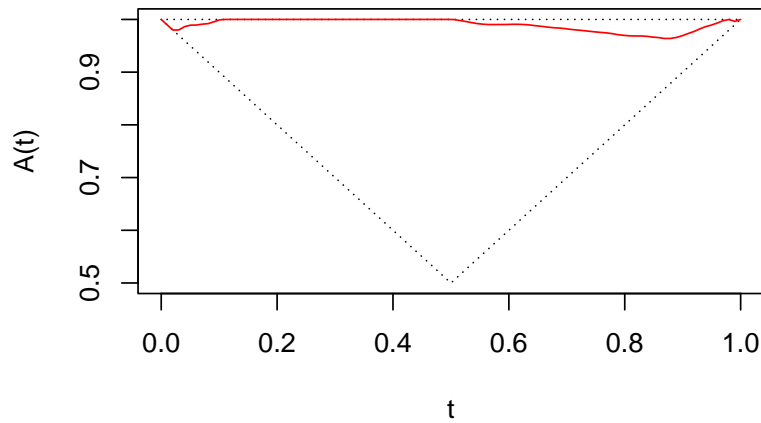


FIGURE 7 – Représentation graphique de l'estimation de la fonction de dépendance de Pickands pour les stations SA et SB

Maintenant, nous pouvons prédire les quantiles extrêmes conditionnels que nous représentons en figure 8. Pour ce faire, nous ajustons les modèles logistique, logistique asymétrique et Hüsler-Reiss sur nos données afin de retenir celui dont l' AIC est minimal. Ici, le modèle retenu est celui de Hüsler-Reiss avec $AIC = 217.724$.

Lorsque p diminue, nous constatons que la valeur des quantiles extrêmes conditionnels augmente de façon exponentielle.

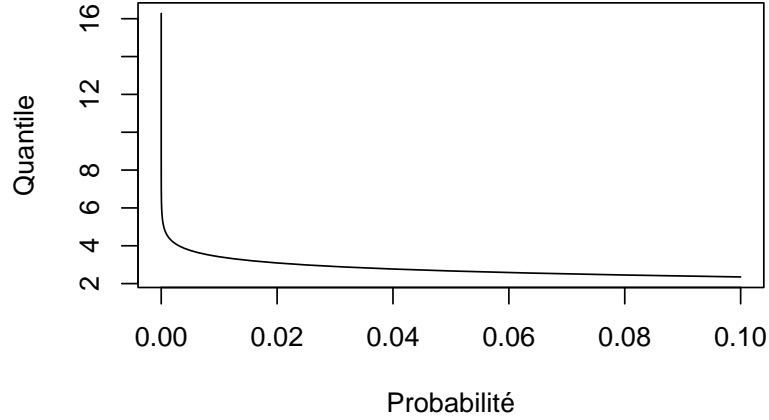


FIGURE 8 – Quantiles conditionnels de SB sachant $SA > 4$ pour une probabilité p allant de 10^{-1} à 10^{-10}

2.3 Analyse des stations SA et 1 (SC)

Afin d'effectuer une seconde analyse bivariable, nous nous intéressons à la station 1, notée SC , qui est plus éloignée de SA que SB avec une distance d'environ 220 kilomètres de SA . Il est important de remarquer que cette station se situe au large des côtes comparativement à l'ensemble des autres stations étudiées.

De la même façon que ci-avant, nous étudions la dépendance entre les données de la station SA et celles de la station SC par la représentation graphique de la fonction de dépendance de Pickands de la figure 16.

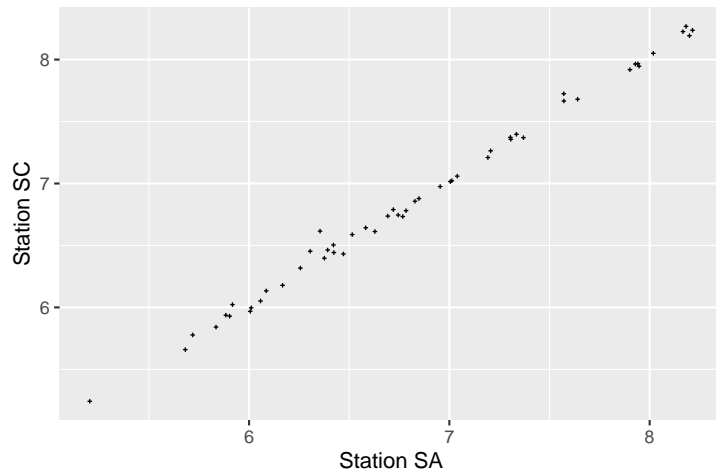


FIGURE 9 – Hauteur maximale des vagues mesurées aux stations SA et SC

En nous basant sur la figure 11, nous supposons que nous sommes dans un cas de dépendance asymptotique voire d'égalité des marginales. Nous testons cette conjecture de la même manière que précédemment, à savoir en calculant χ et $\bar{\chi}$ puis en estimant l'indice extrémal. Nous obtenons $\bar{\chi} = \lim_{u \rightarrow 1} \bar{\chi}(u) = 1$ et $\chi = \lim_{u \rightarrow 1} \chi(u) > 0$. De plus, l'estimateur de θ est $\hat{\theta} \approx 1.03$. Par conséquent, nous sommes bien dans une situation d'égalité des marginales. Nous calculons alors les quantiles conditionnels en tenant compte de cette conclusion.

Pour cette station, nous ajustons à nouveau un modèle de Hüsler-Reiss en procédant par ajustement des trois modèles puis élimination en fonction des valeurs d' AIC obtenues. En prenant $y = 4$, nous obtenons les quantiles conditionnels représentés en figure 10 sur laquelle nous constatons une augmentation des quantiles quand la probabilité p diminue : cette croissance est plus rapide que pour la station SB de la figure 8.

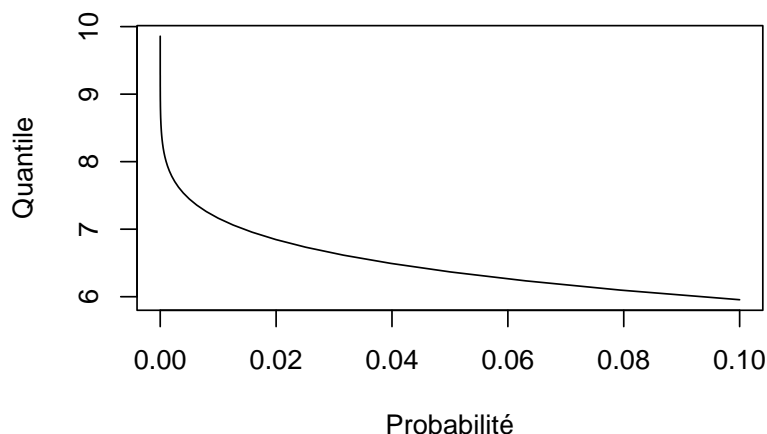


FIGURE 10 – Quantiles conditionnels de SC sachant $SA > 4$ pour une probabilité p allant de 10^{-1} à 10^{-10}

À la station SC , il est plus probable d'atteindre des hauteurs de vagues extrêmes qu'à la station SB sachant que $SA > 4$.

2.4 Analyse des stations SA et 4 (SD)

La dernière analyse bivariable que nous allons effectuer est celle du couple $(SA, 4)$ où la station 4 sera notée SD . Cette dernière est plus proche de SA que SB avec une distance d'environ 2 kilomètres.

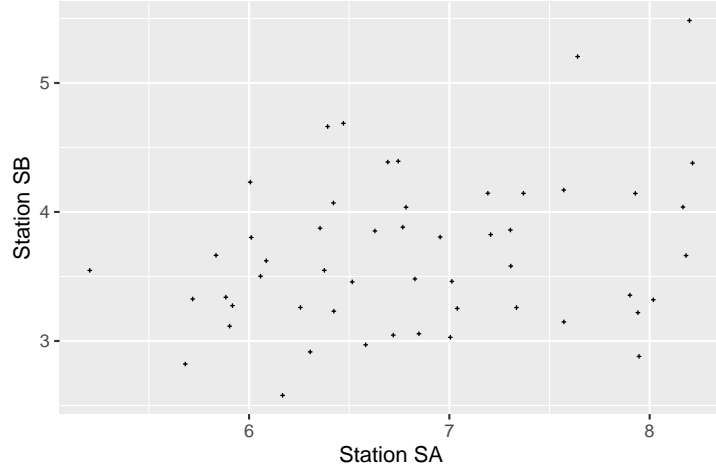


FIGURE 11 – Hauteur maximale des vagues mesurées aux stations SA et SD

À nouveau, nous sommes dans un cas d'indépendance asymptotique avec $\hat{\theta} \approx 1.85$. Nous ajustons encore un modèle de Hüsler-Reiss et calculons les résidus conditionnels, toujours avec $y = 4$: ceux-ci sont représentés en figure 12.

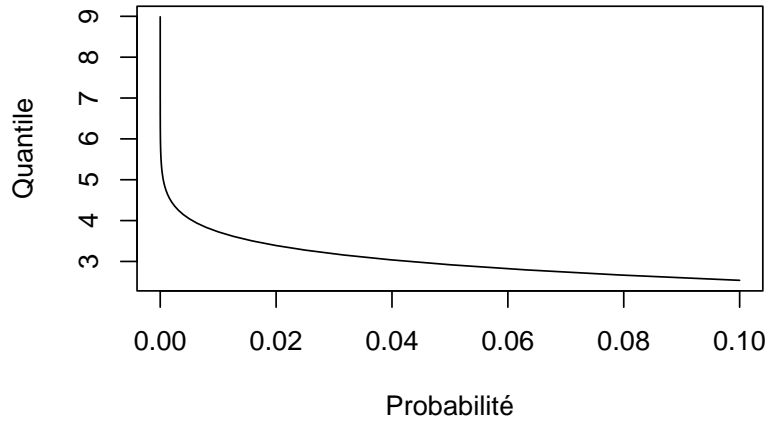


FIGURE 12 – Quantiles conditionnels de SD sachant $SA > 4$ pour une probabilité p allant de 10^{-1} à 10^{-10}

Comme pour les stations SB et SC , nous remarquons que ces quantiles conditionnels augmentent quand p diminue. Cependant, la décroissance des quantiles est plus semblable à celle de la station SB qui est plus rapide que celle de la station SC lorsque la probabilité croît. Par conséquent, la présence de hauteurs de vagues extrêmes de la station SD est similaire à celle de la station SB conditionnellement à la station SA .

2.5 Conclusion

D’abord, nous remarquons qu’en considérant les stations les plus proches de SA , à savoir SB et SD , nous sommes dans une situation d’indépendance asymptotique, ce qui semble paradoxal. Cependant, étant très proches des côtes, le relief autour de ces deux stations peut différer et influencer différemment la hauteur maximale des vagues extrêmes. À l’inverse, nous sommes dans un cas de dépendance asymptotique en considérant la station SC qui est la plus éloignée de SA . il est primordial de noter la localisation très particulière de la station 1 (Annexe. Figure 13). En effet, contrairement aux autres stations, celle-ci se situe au large des côtes ce qui influence fortement les hauteurs de vagues extrêmes mesurées. Nous pouvons supposer que si des vagues particulièrement hautes se forment au large alors cela se ressentira sur les vagues mesurées près des côtes puisque le niveau devient de moins en moins profond et que l’énergie est bloquée par le fond poussant de plus en plus l’eau vers le haut.

Enfin, en cas d’égalité des distributions marginales des couples de stations, les quantiles extrêmes augmentent de manière plus marquée à mesure que la probabilité p diminue, comparativement à une situation d’indépendance asymptotique. La valeur des quantiles dépend directement des valeurs maximales observées dans les différentes stations. Cette disparité s’explique en premier lieu par la méthode d’estimation des quantiles. La force de la dépendance entre les stations peut également contribuer à expliquer les variations dans l’augmentation des quantiles extrêmes. Par exemple, dans le cas d’une dépendance forte entre les stations, les valeurs extrêmes de hauteurs de vagues d’une station peuvent influencer plus fortement celles de l’autre station ce qui peut accélérer l’augmentation des quantiles extrêmes lorsque la probabilité p diminue : ceci serait inversement vrai en cas de dépendance faible.

Annexe

Vous trouverez l'ensemble du code source R utilisé dans ce rapport dans le dépôt Git suivant :
https://github.com/Thamara97/Stat_extremes.



FIGURE 13 – Localisation des stations étudiées dans le Golfe du Lion

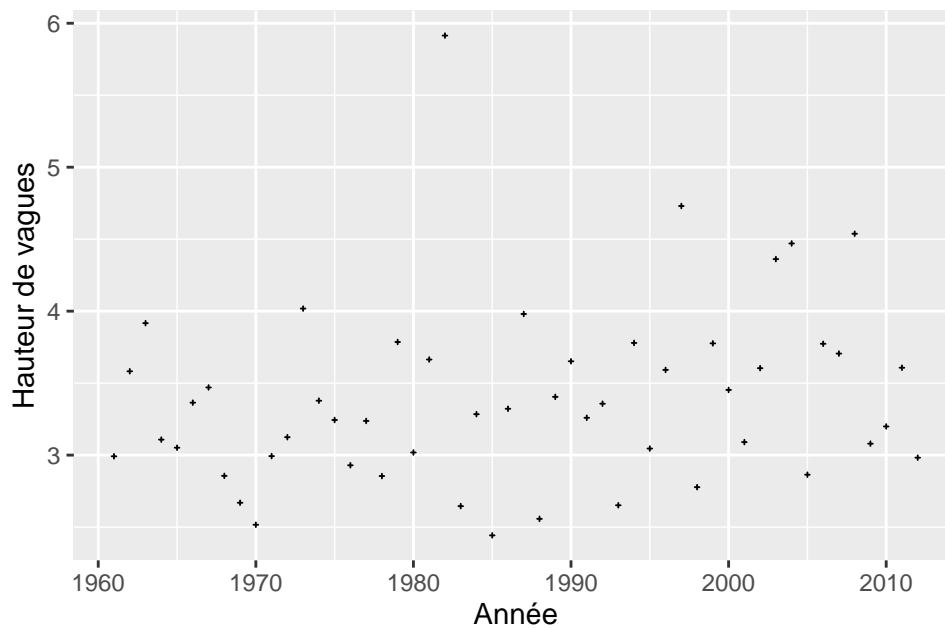


FIGURE 14 – Maxima annuels de hauteur de vagues de la station 12 (SB) du golfe du Lion

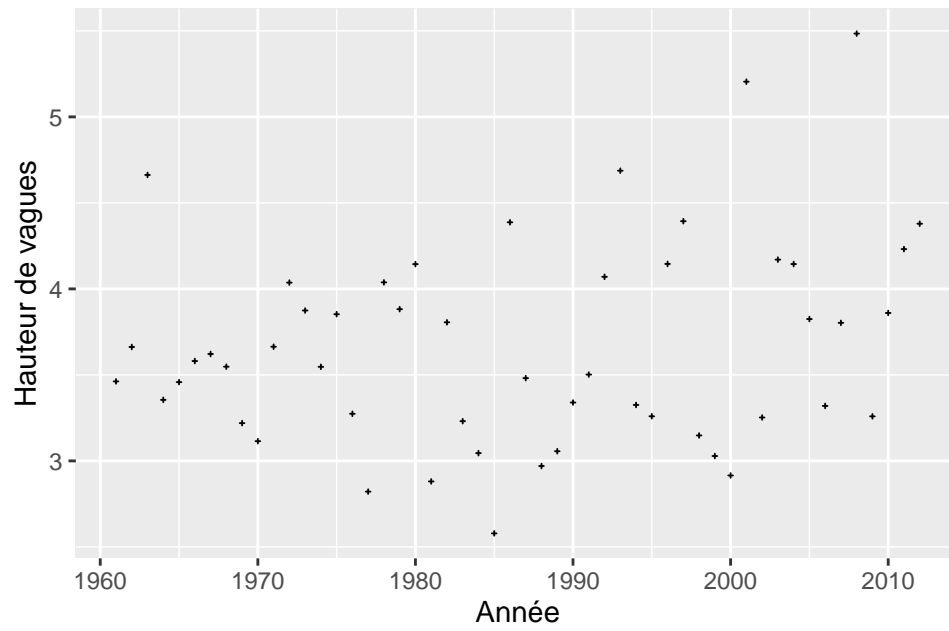


FIGURE 15 – Maxima annuels de hauteur de vagues de la station 4 (*SD*) du golfe du Lion

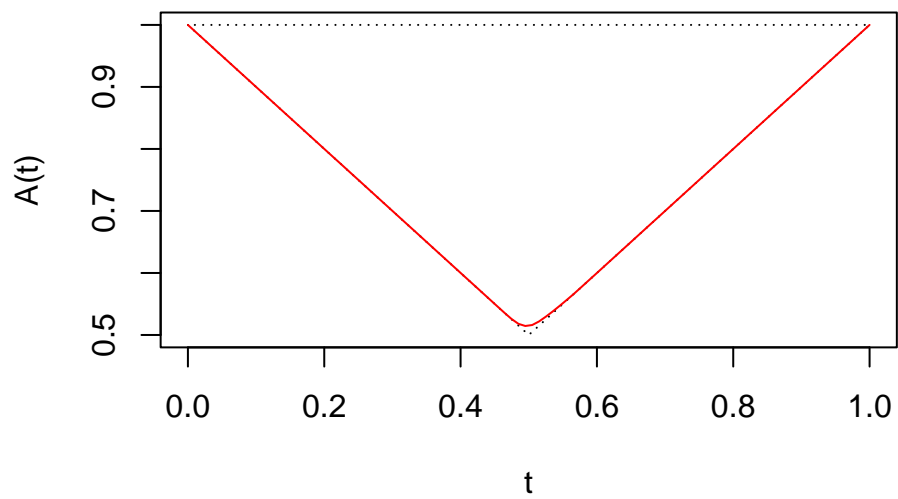


FIGURE 16 – Représentation graphique de l'estimation de la fonction de dépendance de Pickands pour les stations *SA* et *SC*

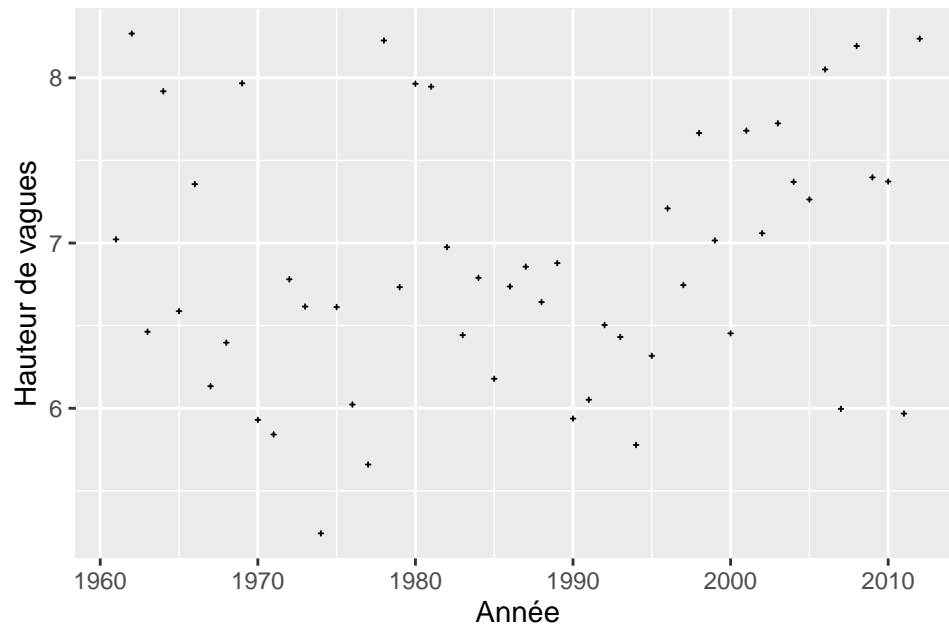


FIGURE 17 – Maxima annuels de hauteur de vagues de la station 1 (*SC*) du golfe du Lion

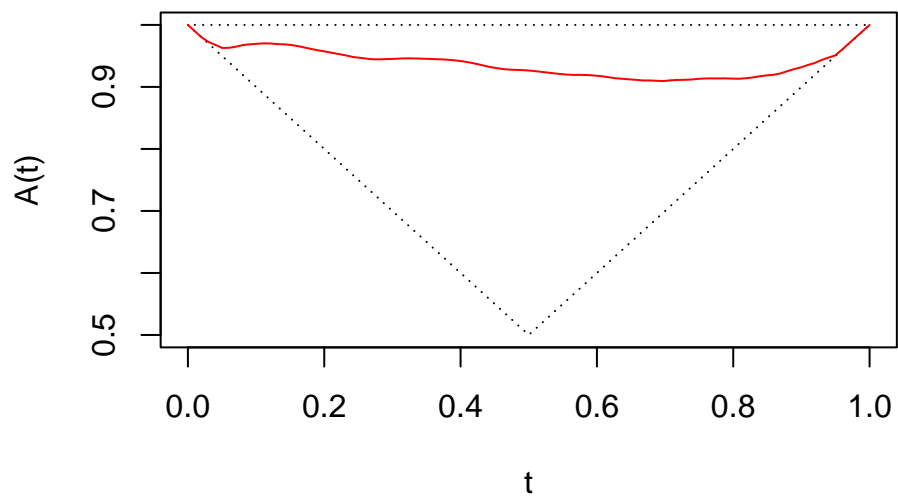


FIGURE 18 – Représentation graphique de l'estimation de la fonction de dépendance de Pickands pour les stations *SA* et *SD*