

# Projet : Vagues dans le golfe du Lion

Damien MARIAC

Lucien DUGOU

14 février 2026

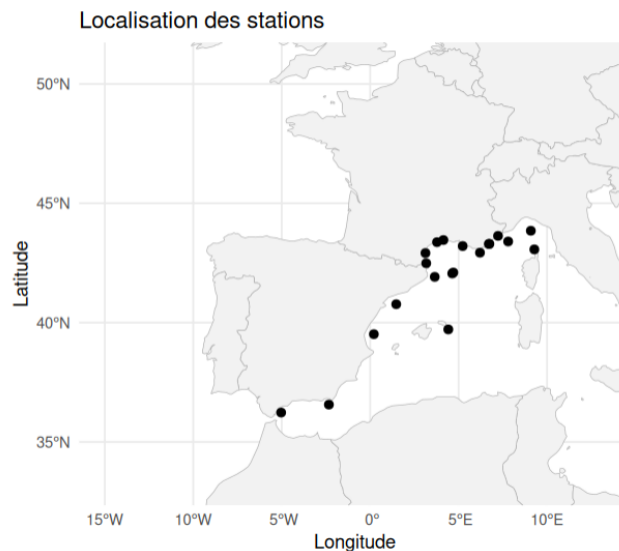
# Table des matières

<b>1</b>	<b>Introduction</b>	<b>3</b>
<b>2</b>	<b>Approche univariée</b>	<b>3</b>
2.1	Maximum par bloc . . . . .	3
2.1.1	Ajustement d'une GEV . . . . .	3
2.1.2	Période de retour . . . . .	5
2.2	Dépassement de seuil . . . . .	5
<b>3</b>	<b>Approche bivariée pour 2 stations choisies</b>	<b>6</b>
3.1	Ajustement d'une distribution bivariée . . . . .	6
3.1.1	Ajustement d'un modèle - logistique, asymétrique logistique et Hüsler Reiss . . . . .	6
3.1.2	Paramètre de dépendance $\theta$ . . . . .	8
3.1.3	Paramètre de dépendance $\chi$ . . . . .	8
<b>4</b>	<b>Implémentation d'une Approche bivariée pour 2 stations proches (puis éloignées)</b>	<b>8</b>

# 1 Introduction

L'objectif de ce projet est de modéliser le comportement extrême des vagues dans le golfe du Lion à l'aide de la théorie des valeurs extrêmes étudiée en cours.

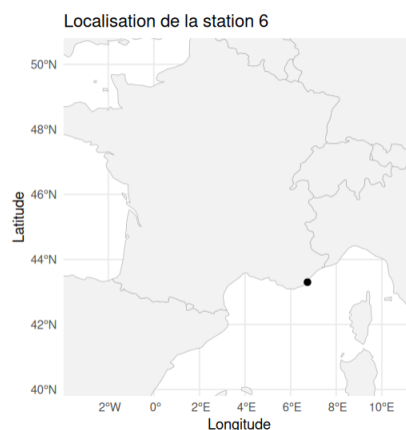
Les données utilisées pour ce projet recensent les hauteurs de vagues de 20 stations différentes de 1961 à 2012 à un rythme de 1 observation par heure.



Nous allons dans une première partie modéliser les hauteurs de vagues extrêmes de la station 6 à l'aide d'une approche univariée. Puis dans une deuxième partie nous allons modéliser la dépendance extrême entre la station 6 et 14 à l'aide d'une approche bivariée. Enfin, dans une troisième partie, nous allons modéliser la dépendance extrême entre des stations proches et éloignées à l'aide d'une approche bivariée.

## 2 Approche univariée

La station 6 est située à Saint Maxime et est représentée sur la carte ci-dessous.

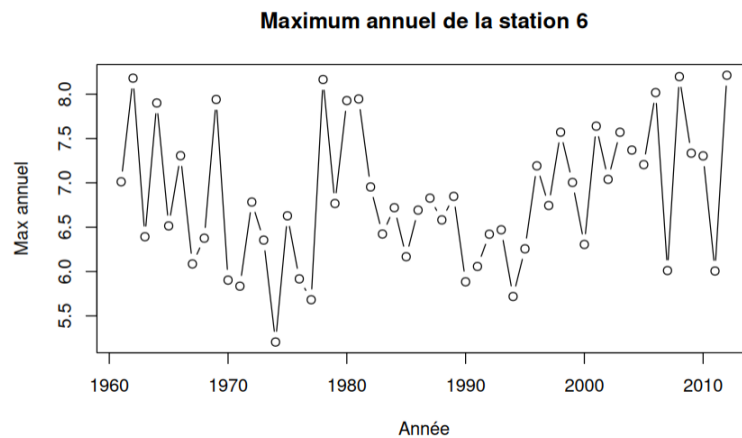


### 2.1 Maximum par bloc

#### 2.1.1 Ajustement d'une GEV

Afin de modéliser les hauteurs de vagues extrêmes, nous allons utiliser la méthode du maximum par bloc. Nous allons diviser les données en blocs annuels et prendre le maximum de chaque bloc. Nous obtenons ainsi une série de 51 maxima annuels.

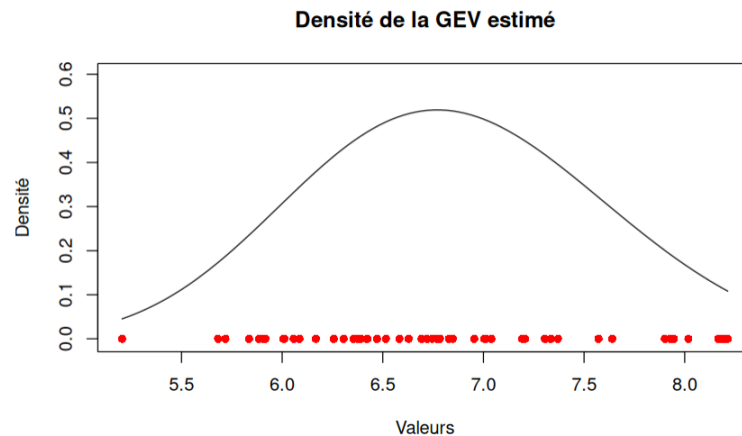
Nous avons choisi de prendre des blocs annuels car cela correspond à une période de temps suffisamment longue pour capturer les variations saisonnières et les tendances à long terme en évitant de la dépendance entre les blocs, tout en étant suffisamment courte pour éviter de perdre des informations importantes sur les événements extrêmes.



Nous pouvons donc fit une distribution de GEV sur ces maxima annuels pour estimer les paramètres de la distribution. Nous récupérons de plus les écarts types de ces paramètres pour pouvoir construire des intervalles de confiance dont nous fixons le niveau à 95%.

- $\mu = 6.560243 \pm 0.2337362$
- $\sigma = 0.7349365 \pm 0.1753661$
- $\gamma = -0.2577396 \pm 0.2722555$

Nous remarquons que le paramètre de forme  $\gamma$  est négatif ce qui correspond à une Weibull, cependant, l'intervalle de confiance contient zéro, ce qui n'exclue pas que la distribution pourrait être de type Gumbel (si  $\gamma = 0$ ) ou de type Fréchet (si  $\gamma > 0$ ).



### 2.1.2 Période de retour

Avec les paramètres estimés, nous pouvons calculer les périodes de retour pour 100, 500 et 1000 ans. Comme nous avons utilisé des maxima annuels, la période de retour est définie par :

$$z_T = \begin{cases} \mu + \frac{\sigma}{\gamma} \left[ \left( -\log \left( 1 - \frac{1}{T} \right) \right)^{-\gamma} - 1 \right] & \text{si } \gamma \neq 0, \\ \mu - \sigma \log \left( -\log \left( 1 - \frac{1}{T} \right) \right) & \text{si } \gamma = 0. \end{cases}$$

Ce qui correspond respectivement à :

- $p_{100} = 8.540445$  m
- $p_{500} = 8.83687$  m
- $p_{1000} = 8.930977$  m

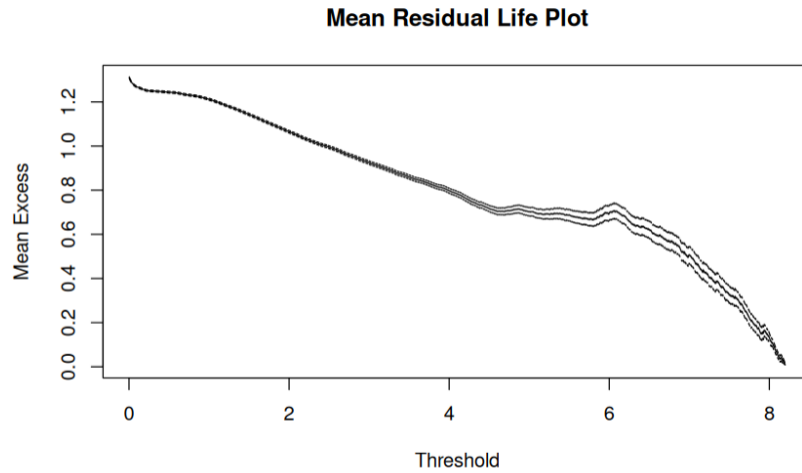
Les périodes de retour semblent relativement proches les unes des autres, ce qui peut s'expliquer par le fait que la distribution de GEV ajustée a une queue bornée (avec un point terminal), ce qui limite la croissance des quantiles extrêmes au fur des années.

Afin de consolider ces résultats, nous pouvons utiliser une autre méthode afin d'estimer les périodes de retour, comme la méthode de dépassement de seuil. En utilisant cette méthode, nous pouvons estimer les périodes de retour pour les mêmes quantiles et comparer les résultats obtenus avec ceux de la méthode du maximum par bloc.

## 2.2 Dépassement de seuil

L'avantage de la méthode de dépassement de seuil est qu'elle utilise plus d'informations que la méthode du maximum par bloc, car elle prend en compte tous les événements qui dépassent un certain seuil, plutôt que de ne considérer que les maxima annuels qui pourraient mettre de côté des valeurs extrêmes. Cependant, le choix du seuil est crucial et peut influencer les résultats obtenus.

Afin de choisir un seuil approprié, nous pouvons utiliser le graphique "mean residual life plot" qui permet d'identifier une zone où les résidus moyens sont linéaires, ce qui suggère que les données suivent une distribution de type GPD au-delà de ce seuil.



En utilisant ce graphique, un seuil de 6m semble être un choix approprié. En effet, à partir de 6 m, le graphique semble présenter une zone approximativement linéaire, compatible avec l'hypothèse GPD.

Après implémentation de la méthode, nous obtenons paramètres avec leur intervalle de confiance à 95% suivants :

- $\sigma = 0.9649 \pm 0.08161178$
- $\gamma = -0.3613 \pm 0.06414328$

Après avoir estimé les paramètres de la GPD, nous pouvons calculer les périodes de retour pour les mêmes quantiles que précédemment. Nous obtenons ainsi :

- $p_{100} = 2.63063$  m donc une hauteur de vague de  $6 + 2.63063 = 8.63063$ m
- $p_{500} = 2.64841$  m donc une hauteur de vague de  $6 + 2.64841 = 8.64841$ m
- $p_{1000} = 2.653404$  m donc une hauteur de vague de  $6 + 2.653404 = 8.653404$ m

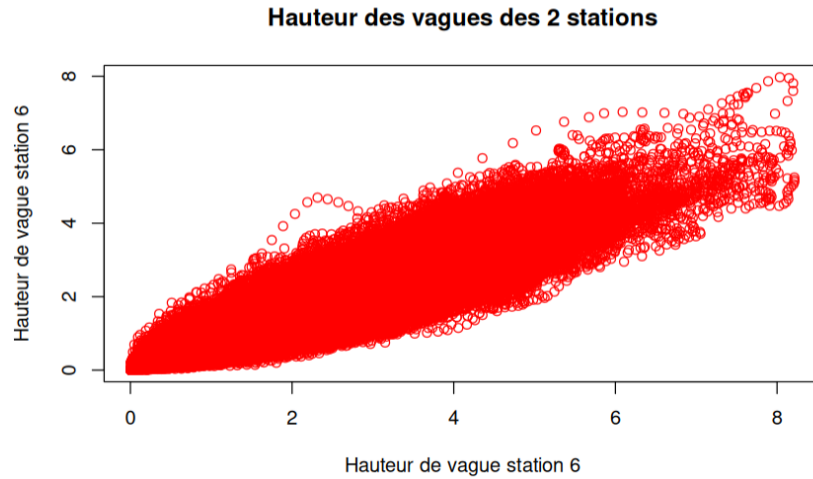
Les périodes de retour obtenues avec la méthode de dépassement de seuil semblent ne pas "exploser" plus le temps avance du au paramètre de forme plus bas que celui de la méthode du maximum par bloc. Néanmoins, les périodes de retour obtenues avec la méthode de dépassement de seuil sont relativement proches de celles obtenues avec la méthode du maximum par bloc, ce qui nous conforte sur la hauteur maximale d'une vague dans les prochaines années.

L'analyse univariée de la station 6 nous a permis d'estimer les paramètres de la distribution de GEV et de GPD, ainsi que les périodes de retour. Nous pouvons approfondir cette analyse en modélisant la dépendance extrême entre la station 6 et une autre station, à l'aide d'une approche bivariable.

### 3 Approche bivariable pour 2 stations choisies

Dans cette section, nous allons modéliser la dépendance extrême entre la station 6 et la station 14, qui est située proche de Sète.

Il semblerait d'après le graphique ci-dessous que les extrêmes des 2 stations aient une dépendance. En effet, on observe que les données avec des grands cordonnées (écarté sur les axes x et y) sont relativement proches de la droite d'équation  $y = x$  ce qui suggère que les extrêmes des 2 stations ont tendance à se produire simultanément.



Afin de fit une distribution bivariable, nous allons transformer les marginales de (distribution de station 6 et de station 14) en distribution de type Frechet. Pour cela, nous appliquons la la fonction de repartition de la GEV avec les parametre estimé sur chaque marginale afin de les transformer en des distributions uniforme, puis nous appliquons la fonction quantile de la distribution de Frechet pour obtenir des données avec des marginales de type Frechet.

Autrement dit, notons  $X_1$  et  $X_2$  les hauteurs de vagues de la station 6 et de la station 14 respectivement,  $F$  la fonction de repartition de la GEV avec les parametres estimé et  $G$  la fonction quantile de la distribution de Frechet, alors nous allons appliquer la transformation suivante :

$$Y_1 = G(F(X_1)) \quad \text{et} \quad Y_2 = G(F(X_2))$$

Et donc  $Y_1$  et  $Y_2$  sont des variables aléatoires avec des marginales de type Frechet.

#### 3.1 Ajustement d'une distribution bivariable

##### 3.1.1 Ajustement d'un modèle - logistique, asymétrique logistique et Hüsler Reiss

Après avoir transformé les données, nous pouvons ajuster un modèle pour modéliser la dépendance extrême entre les deux stations.

### Modèle logistique :

Le modèle logistique est défini par la fonction suivante :

$$G(x, x') = \exp \left( - \left( x^{-1/\alpha} + x'^{-1/\alpha} \right)^\alpha \right)$$

Avec  $\alpha \in ]0, 1]$  le paramètre de dépendance. Lorsque  $\alpha = 1$ , les variables sont indépendantes, et lorsque  $\alpha$  tend vers 0, les variables sont parfaitement dépendantes (même variable).

Après ajustement du modèle logistique, nous obtenons un paramètre de dépendance  $\alpha = 0.7500$ , ce qui suggère une dépendance élevée entre les deux stations. Avec ce modèle, nous obtenons une critère d'AIC de 75167.04.

### Modèle asymétrique logistique :

Le modèle logistique est défini par la fonction suivante :

$$G(x, x') = \exp \left( - (1 - t_1)x^{-1} - (1 - t_2)x'^{-1} - \left( t_1x^{-1/\alpha} + t_2x'^{-1/\alpha} \right)^\alpha \right)$$

Avec  $\alpha \in ]0, 1]$  le paramètre de dépendance,  $t_1 \in [0, 1]$  et  $t_2 \in [0, 1]$  les paramètres d'asymétrie. Lorsque  $t_1 = t_2$ , le modèle asymétrique logistique se réduit au modèle logistique.

Après ajustement du modèle asymétrique logistique, nous obtenons les paramètres suivants :

- $\alpha = 0.6500$
- $t_1 = 0.75$
- $t_2 = 0.75$
- AIC = 74734.89

Avec ce modèle, nous obtenons des valeurs de  $t_1$  et  $t_2$  égales et proches de 1, ce qui suggère que le modèle asymétrique logistique peut se réduire au modèle logistique. De plus, le critère d'AIC du modèle asymétrique logistique est légèrement inférieur à celui du modèle logistique, ce qui suggère que bien que le modèle asymétrique possède plus de paramètres, il n'apporte pas une amélioration significative de l'ajustement par rapport au modèle logistique. Ce dont on aurait pu se douter vu la distribution relativement symétrique des données sur le graphique ci dessus.

### Modèle Hüsler Reiss :

Le modèle Hüsler Reiss est défini par la fonction suivante :

$$V(x, x') = \frac{1}{x} \Phi \left( \frac{\lambda}{2} + \frac{1}{\lambda} \log \left( \frac{x'}{x} \right) \right) + \frac{1}{x'} \Phi \left( \frac{\lambda}{2} + \frac{1}{\lambda} \log \left( \frac{x}{x'} \right) \right)$$

Avec  $\lambda > 0$  le paramètre de dépendance. Lorsque  $\lambda$  tend vers 0, les variables sont parfaitement dépendantes (même variable), et lorsque  $\lambda$  tend vers l'infini, les variables sont indépendantes. Et  $\Phi$  est la fonction de répartition de la loi normale centrée réduite.

Après ajustement du modèle Hüsler Reiss, nous obtenons un paramètre de dépendance  $\lambda$  égal à 1, ce qui suggère la même conclusion que les 2 autres modèles, à savoir une dépendance élevée entre les deux stations. Avec ce modèle, nous obtenons une critère d'AIC de 75641.18 qui est plus élevé que dans le cas du modèle logistique.

### 3.1.2 Paramètre de dépendance $\theta$

Il existe également un paramètre de dépendance  $\theta$  qui permet de mesurer la dépendance des extrêmes entre 2 marginales. En ayant supposé que les marginales suivent une distribution de type Frechet, le paramètre de dépendance  $\theta$  est défini par :

$$G(x, x) = \exp\left(-\theta \frac{1}{x}\right)$$

Où  $G$  est la fonction de répartition de la distribution bivariable ajustée.

Son interprétation est la suivante : lorsque  $\theta = 2$ , les variables sont indépendantes puisque  $G(x, x) = \exp(\frac{-1}{x})\exp(\frac{-1}{x})$  ce qui correspond au produit des marginales. Lorsque  $\theta = 1$ , les variables sont parfaitement dépendantes (même variable) puisque  $G(x, x) = \exp(\frac{-1}{x})$  ce qui correspond bien au max entre  $x$  et  $x$ .

Après ajustement du modèle logistique, nous obtenons un paramètre de dépendance  $\theta$  égal à 1.407525, ce qui suggère une dépendance plus ou moins élevée entre les deux stations, ce qui est cohérent avec les résultats obtenus avec les autres modèles.

### 3.1.3 Paramètre de dépendance $\chi$

Nous pouvons également estimer le paramètre de dépendance  $\chi$  qui est défini par :

$$\chi = \lim_{u \rightarrow 1} P(F_{X_2}(X_2) > u | F_{X_1}(X_1) > u)$$

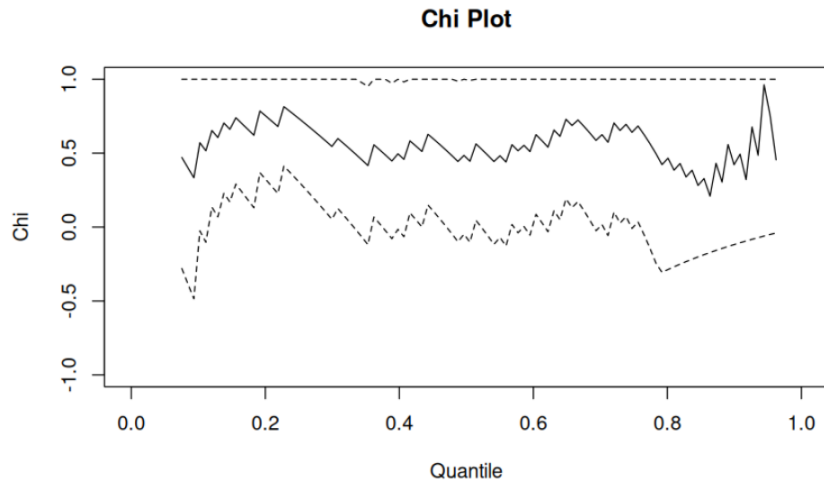
Où  $F_{X_i}$  est la fonction de répartition de la marginale  $X_i$ .

La fonction  $\chi$  peut se réécrire avec un petit jeu de réécriture de la manière suivante :

$$\chi(u) = 2 - \frac{\log C(u, u)}{\log(u)} \rightarrow \chi$$

Où  $C$  est la copule de la distribution bivariable.

Ainsi, en utilisant les données transformées avec des marginales de type Frechet, nous pouvons estimer la copule et calculer  $\chi(u)$  pour différentes valeurs de  $u$  proches de 1. En observant le comportement de  $\chi(u)$  à mesure que  $u$  approche de 1, nous pouvons estimer la valeur limite  $\chi$  qui nous donnera une indication sur la dépendance extrême entre les deux stations.



En observant le graphique ci-dessus, nous pouvons estimer une valeur de  $\chi$  d'environ 0.45, ce qui suggère une dépendance moyenne des extrêmes entre les deux stations. En effet, une valeur de  $\chi$  proche de 1 indique une forte dépendance extrême, tandis qu'une valeur proche de 0 indique l'indépendance des extrêmes.

## 4 Implémentation d'une Approche bivariable pour 2 stations proches (puis éloignées)



**Annexe**