



*Méthode probabiliste pour l'évaluation des phénomènes extrêmes*  
*"hauteurs des marées"*

**Rapport final présenté à la**

**Caisse Centrale de Réassurance**

par la

**Société de Calcul Mathématique SA**

9 mars 2011

## Résumé Opérationnel

Nous présentons la méthode mise en place par la SCM pour estimer la probabilité d'événements rares (observés ou non). Les difficultés proviennent du fait que peu de données sont disponibles ; les méthodes usuelles, pour compenser ces insuffisances, font l'hypothèse de lois paramétriques, comme la loi de Gumbel, la loi exponentielle ou la loi de Poisson.

Notre méthode repose sur les données observées, qui sont discrétisées en classes de manière habituelle (constitution d'un histogramme), et nous ne faisons qu'une seule hypothèse méthodologique : plus le phénomène est extrême, plus sa probabilité est faible. Cette hypothèse est tout à fait raisonnable.

Nous ne faisons pas simplement l'estimation d'une probabilité ; la probabilité de l'événement extrême est elle-même considérée comme une variable aléatoire, que nous appelons "taux de risque" pour éviter la confusion. Nous avons donc une construction de la loi de probabilité du taux de risque, y compris sa variance et ses quantiles.

Le chapitre XVIII du livre "Méthodes probabilistes pour l'étude des phénomènes réels" [1] présente la théorie de la construction de la loi de probabilité du taux de risque, dans le cas d'un phénomène unique (une seule station).

Dans le présent travail, nous étendons cette méthode au cas de plusieurs stations : nous construisons la loi conjointe de plusieurs phénomènes. La théorie n'est pas aussi simple, et il ne suffit pas de dire "plus le phénomène est extrême, plus sa probabilité est faible", comme précédemment, parce qu'ici plusieurs phénomènes peuvent se produire en même temps. On rencontre des difficultés, selon que les stations considérées sont très proches (forte dépendance) ou au contraire éloignées (indépendance). Nous avons retenu les principes suivants :

- Les lois marginales (qui caractérisent chaque station individuellement) doivent avoir la propriété ci-dessus : pour chacune d'elle, plus le phénomène est extrême, plus la probabilité est faible ;
- Pour la loi conjointe, on a une décroissance selon toutes les diagonales : si pour toutes les stations les phénomènes sont plus extrêmes, la probabilité doit décroître.

Techniquement, le calcul de ces probabilités repose sur l'évaluation d'intégrales multiples sur un volume spécifique dans un espace de grande dimension. Dans le présent travail, nous n'avons pas considéré plus de quatre stations en même temps.

Même dans le cas de deux stations, le calcul n'est pas simple. Il y a de nombreux paramètres qui doivent être choisis avec attention, parce que la construction tout entière dépend des choix qui sont faits. Ce sont :

- Les choix des bornes, comme par exemple la valeur maximale de la hauteur des marées : ce choix conditionne la construction, parce qu'il détermine l'extension de la loi de probabilité.

- Les choix de discrétisation : le nombre de classes vides au-delà des résultats observés dépend de la discrétisation qui est faite (taille de la classe). Il faut donc rechercher une valeur optimale.

Pour fixer correctement ces paramètres, nous faisons une étude préliminaire, à propos des particularités régionales : dépendance entre des stations voisines, analyse des événements importants (par exemple : une forte tempête peut n'avoir qu'un caractère local), connaissance d'expert à propos des stations (par exemple, la station de Monaco est mieux "protégée" que celle d'Ajaccio), etc.

Il est intéressant de noter que les dépendances entre stations ne sont pas nécessairement les mêmes pour les phénomènes ordinaires et les phénomènes extrêmes, ces derniers tendant à être plus localisés. C'est en particulier pour cette raison que la discrétisation est importante.

Nous présentons ci-dessous les exemples de calcul de la loi conjointe, pour 1, 2 et 3 stations de la côte méditerranéenne.

### **Cas d'une seule station**

La première chose à faire est de fixer l'intervalle d'extension de la probabilité. La valeur minimum sera le centile à 50 % (médiane) ou l'espérance mathématique ; ce seuil est moins sensible que le quantile à 90 %. Ceci est un choix arbitraire ; il faut fixer une borne inférieure telle que la loi attendue soit décroissante au-delà de cette borne. Dans le cas des hauteurs d'eau, la valeur médiane répond à ces critères et elle est stable (sur les données dont nous disposons).

Pour la valeur maximale, on la fixe en fonction des hauteurs de marées que l'on veut estimer. Cette façon de procéder est assez empirique ; trois règles peuvent être considérées :

- Soit nous ajustons l'extension, pour une station donnée, en fonction de la dernière probabilité non nulle : nous réclamons que la probabilité après ajustement soit la même, pour cette dernière classe.
- Soit nous utilisons comme station de référence une station qui a enregistré les phénomènes les plus extrêmes ; cette station de référence va servir pour toute une région géographiquement homogène. Cela signifie que le pourcentage d'extension observé pour la station de référence va être utilisé ailleurs.
- Soit nous fixons le maximum absolu en fonction d'un avis d'expert.

La taille des classes doit être assez grande, et cohérente avec la discrétisation faite pour les autres stations de la même région : plus la variation de hauteur de marées est grande, plus la taille de la classe doit être élevée.

Pour la station d'Ajaccio, nous obtenons les résultats suivants :

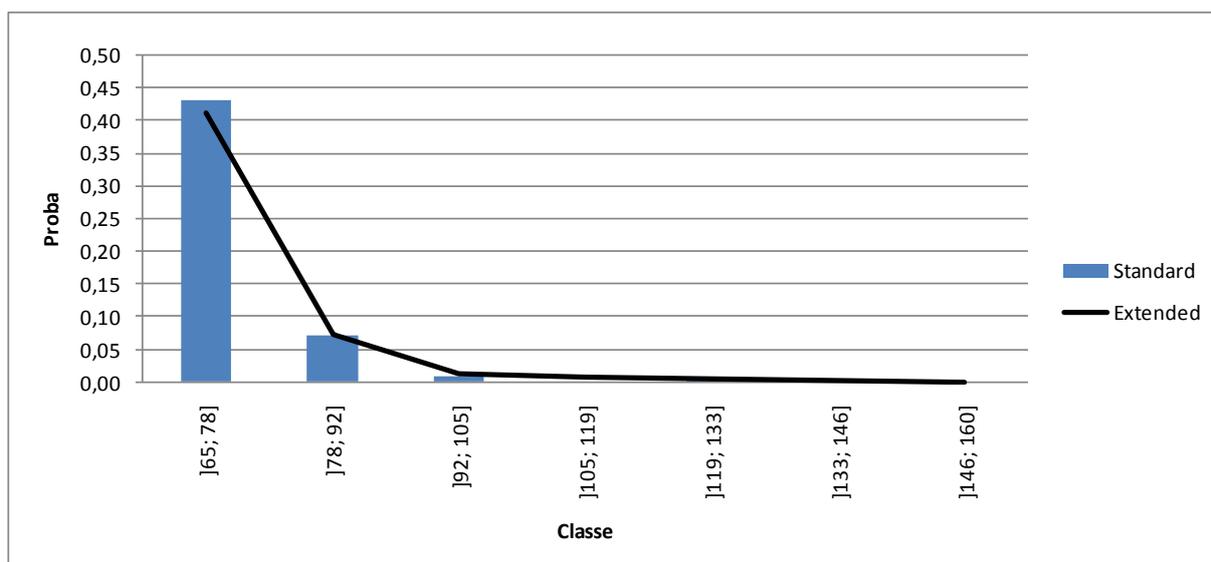


Figure 1 : Lois de probabilité standard et étendue aux phénomènes extrêmes de la station d'Ajaccio

Voici les tables de valeurs pour les probabilités observées et les probabilités calculées :

	Class	Standard	Extended
1	]65; 78]	0,430	0,411
2	]78; 92]	0,072	0,074
3	]92; 105]	0,008	0,014
4	]105; 119]	0	0,007
5	]119; 133]	0,004	0,004
6	]133; 146]	0	0,0023
7	]146; 160]	0	0,0009

Tableau I : Probabilité des classes des lois standard et étendue pour la station d'Ajaccio

La station d'Ajaccio a enregistré les hauteurs de marées les plus importantes. Nous l'utilisons comme station de référence pour la côte méditerranéenne.

Nous obtenons les résultats suivants, pour les stations de Marseille et Monaco :

	Class Marse	Standard	Extended		Class Monac	Standard	Extended
1	]71; 87]	0,373	0,360	1	]69; 83]	0,460	0,446
2	]87; 102]	0,163	0,160	2	]83; 96]	0,087	0,087
3	]102; 118]	0,034	0,038	3	]96; 110]	0,019	0,023
4	]118; 133]	0,015	0,020	4	]110; 123]	0	0,0065
5	]133; 149]	0,011	0,012	5	]123; 137]	0	0,0030
6	]149; 164]	0	0,0044	6	]137; 150]	0	0,0013
7	]164; 180]	0	0,0016				

Tableau II : Probabilité des classes des lois standard et étendue pour les stations de Marseille et Monaco

### Loi conjointe de deux stations

Voici d'abord la loi conjointe standard des stations d'Ajaccio et de Marseille obtenue à partir des données enregistrées :

Ajacc\Marse	1	2	3	4	5	6	7	somme
1	0,1901141	0,1140684	0,0152091	0,0076046	0,0114068	0	0	0,3384
2	0,0152091	0,0380228	0,0152091	0,0038023	0	0	0	0,0722
3	0	0,0038023	0,0038023	0	0	0	0	0,0076
4	0	0	0	0	0	0	0	0
5	0	0	0	0,0038023	0	0	0	0,0038
6	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0
somme	0,2053	0,1559	0,0342	0,0152	0,0114	0	0	

Tableau III : Loi conjointe standard des stations de Marseille et Monaco

Pour construire la loi conjointe étendue, nous construisons un "tableau de Young", avec les propriétés suivantes :

- les valeurs sont décroissantes sur toutes les diagonales ;
- elles sont décroissantes sur chaque ligne et chaque colonne.

Un "tableau de Young" est une matrice de nombres, décroissants selon les lignes et les colonnes. Ici, nous avons besoin de propriétés particulières.

Si les stations sont indépendantes, ce choix de conditions est légitime. Mais dans le cas de la dépendance entre stations, la construction ne sera appliquée qu'à un certain ensemble de valeurs. La construction de la loi conjointe requiert donc une analyse préliminaire de la table des valeurs observées.

Dans le cas d'Ajaccio et de Marseille, nous construisons des tableaux de Young de la manière suivante :

- premier groupe de données : classes 1 à 4 x classes 1 à 4 ;
- deuxième groupe de données : classes 2 à 5 x classes 2 à 5, etc.

Les tableaux de Young ainsi générés correspondent à un masque de taille (4x4) qui est appliqué à l'ensemble de la loi conjointe en quatre étapes. Nous obtenons la table suivante pour la loi conjointe étendue :

Ajacc\Marse	1	2	3	4	5	6	7	somme
1	0,1664381	0,0974985	0,0227162	0,0118744	0,0114068	0	0	0,3099
2	0,0330491	0,01784	0,0099827	0,0057859	0,0031804	0	0	0,0698
3	0,0091464	0,0074008	0,0031943	0,0018995	0,0012305	0,0007907	0	0,0237
4	0,0041779	0,004225	0,0018898	0,0007922	0,0004722	0,0003122	0,000207	0,0121
5	0	0,0025827	0,0012276	0,0004721	0,0002636	0,0001593	0,0000933	0,0048
6	0	0	0,0007891	0,0003123	0,0001593	0,0000858	0,0000417	0,0014
7	0	0	0	0,0002068	0,000093	0,0000417	0,0000143	0,0004
somme	0,2128	0,1295	0,0398	0,0213	0,0168	0,0014	0,0004	

Tableau IV : Loi conjointe étendue aux phénomènes extrêmes des stations de Marseille et Monaco

L'extension à l'ensemble du tableau ne donne pas toujours de bons résultats ; c'est pourquoi nous sommes amenés à introduire des "masques", qui correspondent à des tableaux partiels.

Les graphes des lois marginales sont les suivants :

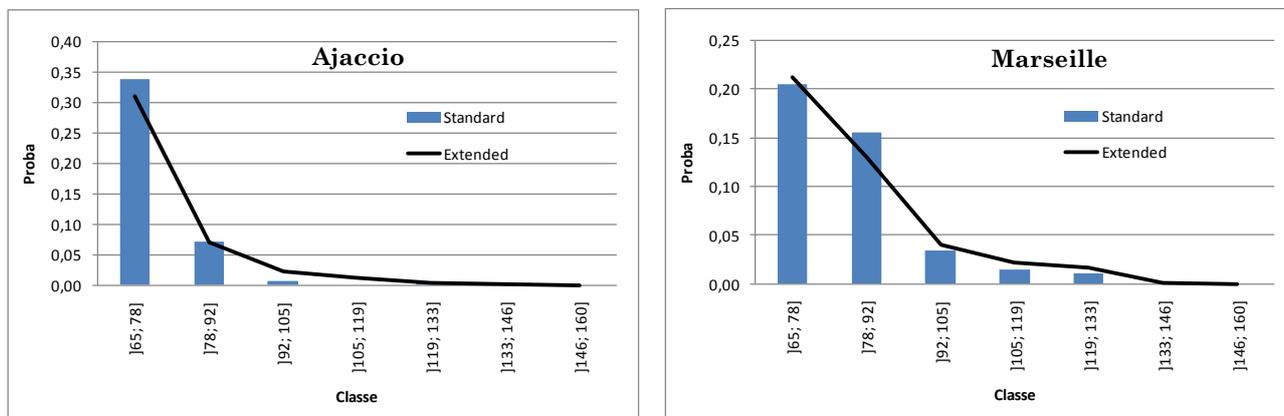


Figure II : Lois conjointes standard et étendue aux phénomènes extrêmes des stations d'Ajaccio et de Marseille

### Loi conjointe de trois stations

Un masque de taille (4x4x3) est appliqué à la loi conjointe des stations d'Ajaccio, Marseille et Monaco.

Les lois marginales standards et étendues sont les suivantes :

Class	Ajaccio		Marseille		Monaco	
	Standard	Extended	Standard	Extended	Standard	Extended
1	0,3118	0,2372	0,1863	0,1766	0,2890	0,2399
2	0,0722	0,0793	0,1483	0,1102	0,0875	0,0955
3	0,0076	0,0457	0,0342	0,0565	0,0190	0,0505
4	0	0,0268	0,0152	0,0342	0	0,0083
5	0,0038	0,0056	0,0114	0,0170	0	0,0011
6	0	0,0007	0	0,0007	0	0,0001
7	0	0,0001	0	0,0001		

Figure V : Lois marginales standards et étendues aux phénomènes extrêmes des stations d'Ajaccio, Marseille et Monaco

La loi conjointe étendue dépend de la taille du masque qui a été appliqué. En comparant, les lois marginales étendues à la loi de probabilité étendue construite pour chaque station, il est possible de contrôler la validité de la loi conjointe obtenue.

Les événements de probabilité non nulle de la loi conjointe étendue aux phénomènes extrêmes des stations d'Ajaccio, Marseille et Monaco sont répertoriés dans le tableau ci-dessous :

Ajacc	Marse	Monac	Standard	Extended		Ajacc	Marse	Monac	Standard	Extended
1	1	1	0,1673004	0,09059		4	4	3	0	0,00024
1	1	2	0,0038023	0,01612		4	4	4	0	0,00007
1	1	3	0	0,00943		4	4	5	0	0,00005
1	2	1	0,0760456	0,04682		4	4	6	0	0,00003
1	2	2	0,0304183	0,01225		4	5	2	0	0,00052
1	2	3	0	0,00653		4	5	3	0	0,00016
1	3	1	0,0076046	0,01567		4	5	4	0	0,00005
1	3	2	0,0076046	0,00710		4	5	5	0	0,00003
1	3	3	0	0,00486		4	5	6	0	0,00002
1	4	1	0,0038023	0,00796		4	6	3	0	0,00011
1	4	2	0,0038023	0,00513		4	6	4	0	0,00003
1	4	3	0	0,00331		4	6	5	0	0,00002
1	5	1	0,0038023	0,00380		4	6	6	0	0,00001
1	5	2	0,0076046	0,00760		4	7	4	0	0,00002
2	1	1	0,0152091	0,01518		4	7	5	0	0,00001
2	1	2	0	0,00969		4	7	6	0	0,00001
2	1	3	0	0,00629		5	2	2	0	0,00153
2	2	1	0,0114068	0,01124		5	2	3	0	0,00084
2	2	2	0,0228137	0,00520		5	2	4	0	0,00045
2	2	3	0,0038023	0,00313		5	3	2	0	0,00089
2	2	4	0	0,00192		5	3	3	0	0,00027
2	3	1	0,0038023	0,00865		5	3	4	0	0,00016
2	3	2	0,0038023	0,00333		5	3	5	0	0,00010
2	3	3	0,0076046	0,00197		5	4	2	0	0,00055
2	3	4	0	0,00113		5	4	3	0,0038023	0,00016
2	4	1	0	0,00442		5	4	4	0	4,64E-05
2	4	2	0,0038023	0,00230		5	4	5	0	2,86E-05
2	4	3	0	0,00132		5	4	6	0	1,93E-05
2	4	4	0	0,00072		5	5	2	0	3,09E-04
2	5	2	0	0,00156		5	5	3	0	1,03E-04
2	5	3	0	0,00084		5	5	4	0	2,88E-05
2	5	4	0	0,00045		5	5	5	0	1,65E-05
3	1	1	0	0,00866		5	5	6	0	1,02E-05
3	1	2	0	0,00611		5	6	3	0	6,62E-05
3	1	3	0	0,00312		5	6	4	0	1,96E-05
3	2	1	0	0,00659		5	6	5	0	1,03E-05
3	2	2	0	0,00322		5	6	6	0	5,78E-06
3	2	3	0,0038023	0,00191		5	7	4	0	1,37E-05
3	2	4	0	0,00111		5	7	5	0	6,47E-06
3	3	1	0	0,00427		5	7	6	0	3,20E-06
3	3	2	0,0038023	0,00205		6	3	3	0	2,00E-04
3	3	3	0	0,00061		6	3	4	0	1,10E-04
3	3	4	0	0,00038		6	3	5	0	6,46E-05
3	3	5	0	0,00026		6	4	3	0	1,11E-04
3	4	1	0	0,00352		6	4	4	0	3,33E-05
3	4	2	0	0,00136		6	4	5	0	1,95E-05
3	4	3	0	0,00038		6	4	6	0	1,24E-05
3	4	4	0	0,00023		6	5	3	0	6,63E-05
3	4	5	0	0,00016		6	5	4	0	1,96E-05
3	5	2	0	0,00087		6	5	5	0	1,04E-05
3	5	3	0	0,00027		6	5	6	0	5,79E-06
3	5	4	0	0,00016		6	6	3	0	3,94E-05
3	5	5	0	0,00010		6	6	4	0	1,26E-05
3	6	3	0	0,00020		6	6	5	0	5,85E-06
3	6	4	0	0,00011		6	6	6	0	2,79E-06
3	6	5	0	0,00006		6	7	4	0	8,17E-06
4	1	1	0	0,00569		6	7	5	0	3,21E-06
4	1	2	0	0,00343		6	7	6	0	1,22E-06
4	1	3	0	0,00232		7	4	4	0	2,46E-05
4	2	1	0	0,00331		7	4	5	0	1,36E-05
4	2	2	0	0,00220		7	4	6	0	8,11E-06
4	2	3	0	0,00127		7	5	4	0	1,37E-05
4	2	4	0	0,00070		7	5	5	0	6,47E-06
4	3	1	0	0,00234		7	5	6	0	3,20E-06
4	3	2	0	0,00135		7	6	4	0	8,17E-06
4	3	3	0	0,00038		7	6	5	0	3,21E-06
4	3	4	0	0,00023		7	6	6	0	1,22E-06
4	3	5	0	0,00016		7	7	4	0	4,87E-06
4	4	1	0	0,00115		7	7	5	0	1,48E-06
4	4	2	0	0,00086		7	7	6	0	3,43E-07

Tableau VI : Loi conjointe étendue aux phénomènes extrêmes des stations d'Ajaccio, Marseille et Monaco

## ***Traitement et sélection des données***

L'étude menée porte sur les hauteurs d'eau côtières mesurées par le SHOM le long des côtes de la Méditerranée, de la Manche et de l'Atlantique. Nous disposions au début de l'étude des hauteurs d'eau horaires des ports suivants :

- Ajaccio ;
- Boucau ;
- Boulogne-sur-mer ;
- Brest ;
- Calais ;
- Cherbourg ;
- Concarneau ;
- Dieppe ;
- Dunkerque ;
- Arcachon-Eyrac ;
- Le Havre ;
- Le Conquet ;
- Le Crouesty ;
- La Rochelle ;
- Marseille ;
- Monaco ;
- Nice ;
- Les Sables d'Olonne ;
- Port-Bloc ;
- Port-Vendres ;
- Roscoff ;
- Sète ;
- Saint Jean-de-Luz ;
- Saint Malo ;
- Saint Nazaire ;
- Toulon.

Dans un premier temps, nous avons scindé les stations entre celles se situant le long de la côte méditerranéenne et celles se situant le long de l'Atlantique et de la Manche.

En raison de la quantité très importante de données manquantes, il est nécessaire d'utiliser un niveau de précision moins fin dans le but de réduire la part des données à reconstituer. L'étude porte donc sur les hauteurs d'eau maximales mensuelles qui sont déduites des données horaires.

Certaines stations présentant des historiques trop courts, elles seront ignorées dans l'étude. Afin de réduire encore la quantité de données manquantes, l'étendue de l'historique à utiliser pour chacune des côtes est définie au préalable.

Ainsi pour la Méditerranée, l'étude porte sur les 5 stations suivantes : Ajaccio, Monaco, Marseille, Nice et Toulon. L'historique considéré est de 263 mois (environ 22 ans) répartis sur deux périodes distinctes : d'août 1981 à juillet 1985 puis d'octobre 1991 à juin 2010.

Pour la côte Atlantique et la Manche, 12 stations ont été sélectionnées :

- Boucau ;
- Boulogne-sur-mer ;
- Brest ;
- Cherbourg ;
- Dieppe ;
- Le Havre ;
- Le Conquet ;
- La Rochelle ;
- Les Sables d'Olonne ;
- Roscoff ;
- Saint Jean-de-Luz ;
- Saint Nazaire.

L'historique étudié pour ces stations s'étend sur 474 mois consécutifs (environ 39 ans), allant de janvier 1971 à juin 2010.

La possibilité d'exploiter les données HYDRO concernant les hauteurs d'eau des fleuves côtiers avait été considérée au début de notre étude. Cependant, l'analyse de ces données a montré que les mesures dont nous disposons actuellement ne sont pas satisfaisantes :

- Soit aucune station SHOM ne se trouve proche de l'embouche des fleuves côtiers des stations HYDRO ;
- Soit les stations HYDRO sont situées trop loin de la côte.

Il n'est donc pas pertinent d'utiliser ces données, car aucune relation avec des phénomènes extrêmes (tempête) n'a pu être mise en évidence.

## Table des matières

I.	Contexte et origine des données utilisées .....	12
A.	Objectifs de l'étude .....	12
B.	Zone de travail et origine des données.....	12
1.	Hauteurs d'eau côtières.....	12
2.	Hauteurs d'eau des cours d'eau côtiers.....	13
II.	Données SHOM.....	14
A.	Collecte des données SHOM .....	14
B.	Sélection et mise en forme des données SHOM.....	14
C.	Première analyse statistique : disponibilité des données horaires .....	16
D.	Analyse statistique après sélection des hauteurs maximales mensuelles.....	18
E.	Etude de la corrélation entre les stations SHOM.....	20
F.	Détection d'évènements extrêmes .....	24
G.	Prise en compte de la houle : croisement des données avec celles de l'historique CMV de la CCR .....	29
1.	Première étude .....	30
2.	Seconde approche : étude de la station BOUCAU.....	32
H.	Corrélation des stations SHOM sur les phénomènes extrêmes .....	33
1.	Corrélation des phénomènes extrêmes .....	33
2.	Corrélation des phénomènes extrêmes avec les autres événements.....	37
III.	Données HYDRO .....	41
A.	Sélection des données HYDRO .....	41
B.	Collecte et mise en forme des données HYDRO .....	41
C.	Disponibilité des données HYDRO .....	42
D.	Etude de la corrélation aux données SHOM .....	44
1.	Emplacement des stations SHOM et HYDRO.....	44
2.	Corrélation sur tout l'historique .....	45
E.	Etude de la corrélation des évènements extrêmes .....	48
1.	Résultats en Méditerranée.....	48
2.	Résultats sur la côte Atlantique et la Manche .....	51
F.	Corrélation entre la station SHOM de Cherbourg et la station HYDRO I6943110 (La Divette).....	56

IV.	Sélection des stations exploitables .....	57
A.	Critères de sélection des stations.....	57
B.	Exemple de sélection de deux stations .....	57
1.	Choix du littoral .....	57
2.	Choix de l'historique.....	57
3.	Reconstruction des données manquantes .....	58
C.	Stations exploitables .....	59
V.	Extension de la loi conjointe : théorie.....	60
A.	Premiers éléments.....	60
B.	Le simplexe de travail.....	61
C.	Applications .....	62
D.	Calcul de l'intégrale.....	63
1.	Etape 1.....	63
2.	Etape 2.....	64
3.	Génération de Tableaux de Young - algorithmes .....	65
4.	Description fonctionnelle de l'algorithme : .....	67
VI.	Extension de la loi conjointe : mise en œuvre pratique .....	70
A.	Les difficultés .....	70
1.	Cas de la station de Brest .....	70
2.	Cas de la station de La Rochelle .....	71
3.	Conclusion .....	71
B.	Construction de la loi de probabilité standard étendue. ....	72
1.	Application à la station d'Ajaccio.....	72
	Etape 1: discrétisation en classes.....	74
	Etape 2 : Initialisation du taux de risque .....	74
	Etape 3 : Calcul de la loi de probabilité de l'ensemble $(p_1, \dots, p_K)$ :.....	75
	Etape 4 : Estimation de la loi marginale :.....	75
2.	Application à la station de Monaco.....	78
3.	Application à la station de Marseille .....	79
C.	Construction de la loi conjointe étendue pour deux stations .....	80
	Etape 1: discrétisation en classes :.....	81
	Etape 2 : Initialisation du taux de risque .....	82

# I. Contexte et origine des données utilisées

## A. Objectifs de l'étude

La Caisse Centrale de Réassurance (CCR) intervient dans la réassurance des sinistres importants, notamment en cas de catastrophe naturelle (inondations, etc.). Ce sont, par définition, des événements rares et de fortes conséquences.

L'étendue de chaque événement et sa probabilité vont conditionner la prime d'assurance correspondante, mais ni l'étendue ni la probabilité ne sont faciles à définir. Dans le cas d'une inondation, par exemple, cela signifie : quelle sera la zone couverte, quelle sera la hauteur d'eau, et combien de fois par siècle peut-on attendre un phénomène d'ampleur donnée ?

On dispose généralement d'une information locale (par exemple la hauteur de la marée en un certain nombre de points, pendant un certain nombre d'années), mais il est délicat d'en déduire une information globale portant sur une zone entière.

La CCR s'intéresse aux hauteurs d'eau côtières, et en particulier aux phénomènes extrêmes : ceux qui ont été rarement, voire jamais observés. Ces événements peuvent provoquer des submersions côtières et donc des dégâts importants. Il est nécessaire pour la CCR d'évaluer la probabilité de tels événements, non pas ponctuellement en chaque station de mesure, mais globalement, sur l'ensemble des zones surveillées.

Il s'agit ici de construire la loi de probabilité conjointe des hauteurs d'eau (hauteur de la marée) en un certain nombre de stations de mesures sur le littoral de la France métropolitaine. Nous mettrons ensuite en place une méthode permettant de reconstituer la loi de probabilité des phénomènes extrêmes sur l'ensemble de ces stations.

## B. Zone de travail et origine des données

Le travail porte sur les hauteurs côtières, sur l'ensemble du littoral de la France métropolitaine. Ces données pourront être complétées par l'historique des hauteurs d'eau des cours d'eau côtiers, et éventuellement de données de houle disponibles auprès du CETMEF (Centre d'Etudes Techniques Maritimes Et Fluviales).

### 1. Hauteurs d'eau côtières

Le SHOM (Service Hydrographique et Océanographique de la Marine) dispose d'un certain nombre de marégraphes répartis sur le littoral français (métropole et Outre-Mer). Cet organisme met à disposition, sur le site [www.soneil.org/ftpWeb/index.php](http://www.soneil.org/ftpWeb/index.php), l'historique des hauteurs d'eau de 31 stations. Plusieurs types de données sont disponibles :

- Les hauteurs d'eau (en cm) sur les 24 créneaux horaires de chaque jour de l'année, pour un certain nombre d'années, propre à chaque station ;

- Les hauteurs par créneau de 10 minutes de chaque jour de l'année, pour un certain nombre d'années par station ;
- Les niveaux moyens par jour, mois et année : il s'agit de la moyenne de l'écart entre la hauteur observée à un instant donné et la marée calculée au même instant (surcote).

## 2. Hauteurs d'eau des cours d'eau côtiers

Sur chaque cours d'eau en France, des stations de mesures récoltent des données de hauteur d'eau.

Ces mesures sont effectuées par différents services de l'État comme les DIREN, les Agences de l'Eau, les Services de Prévision des Crues, le Cemagref... Chaque producteur de données vérifie et valide ses mesures selon sa propre appréciation, puis alimente la « Banque HYDRO ».

HYDRO calcule sur une station donnée les débits instantanés, journaliers, mensuels... à partir des valeurs de hauteur d'eau et des courbes de tarage (relations entre les hauteurs et les débits). Ces valeurs sont actualisées à chaque mise à jour d'une hauteur ou d'une courbe de tarage (addition, précision supplémentaire, correction...).

Nous avons ainsi accès à quatre types de données :

- *HMM* : hauteur d'eau maximale instantanée mensuelle (en mètres et centimètres) ;
- *QJO* : débit journalier correspondant à la moyenne des débits sur une journée donnée (en litres par seconde et mètres cubes par seconde) ;
- *QME* : débit mensuel observé (en litres par seconde et mètres cubes par seconde) ;
- *QMM* : débit mensuel maximum (en litres par seconde et mètres cubes par seconde) ;

## II. Données SHOM

### A. Collecte des données SHOM

La première étape du travail a consisté à télécharger les données du site [www.sonel.org/ftpWeb/index.php](http://www.sonel.org/ftpWeb/index.php). Les données peuvent être téléchargées :

- Soit en utilisant un navigateur internet classique, ce qui est laborieux, dans la mesure où le site Internet impose une limite à la quantité de documents téléchargés quotidiennement ;
- Soit à l'aide d'un client FTP (ex : FileZilla), ce qui permet de récupérer les données voulues rapidement.

Nous avons sélectionné les marégraphes situés en France métropolitaine : nous disposons des données relevées en 26 stations de mesure, dont la localisation est donnée dans la carte ci-dessous (réalisée sous Arcgis).



Figure 1 : Localisation des marégraphes en France métropolitaine

### B. Sélection et mise en forme des données SHOM

Nous choisissons dans un premier temps d'utiliser les relevés horaires de hauteurs d'eau : le niveau de précision des créneaux de 10 minutes n'est pas nécessaire dans cette étude.

La taille de l'historique varie fortement selon la station considérée. De plus, les historiques des bases horaires et par créneaux de 10 minutes peuvent être disjoints : certaines années

sont présentes dans les relevés toutes les 10 minutes et pas dans les relevés horaires, et inversement.

Nous avons donc constitué une base unique de relevés horaires des 26 stations, à partir des deux bases de données :

- Lorsqu'elles sont disponibles, les données horaires sont issues de la base de relevés horaires ;
- Lorsqu'une date est présente dans les relevés par créneaux de 10 minutes et pas dans la base horaire, nous retenons le maximum des 6 hauteurs relevées dans l'heure.

Les données téléchargées comportent, pour chaque station :

- Un champ comportant le code station ;
- Deux champs relatifs à la date : année et jour de l'année (par exemple 1981 et 151 pour le 151<sup>ème</sup> jour de l'année 1981) ;
- 24 champs comportant les 24 hauteurs horaires relevées ce jour par le marégraphe de la station.

Nous avons fusionné l'ensemble des données sous une forme matricielle, sous Excel :

- Une ligne par triplet année, jour de l'année (1 à 365, 366 pour les années bissextiles), et créneau horaire (1 à 24). Les trois premiers champs contiennent ces informations ;
- L'historique de chaque station est contenu dans une colonne unique, avec pour en-tête le code station.

En raison des limites imposées par Excel en termes de nombre de lignes d'un tableau, l'historique ainsi fusionné a été scindé en deux tableaux de mêmes dimensions (721 716 x 29 chacun).

Ceci permet une approche globale, ce qui est tout à fait adapté à la construction de la loi conjointe. De plus, cette mise en forme est similaire à celle utilisée dans le cadre de notre contrat précédent (construction de la loi conjointe des débits de rivières du bassin versant de la Seine). Nous pourrions ainsi adapter facilement les algorithmes existants. Pour cette même raison, nous choisissons de conserver les données au format Excel.

### C. Première analyse statistique : disponibilité des données horaires

Face à la variabilité de la taille de l'historique parmi les 26 stations, il est nécessaire d'étudier la disponibilité des relevés.

Le tableau suivant contient, pour chaque station, les dates de début et fin d'historique, le nombre de relevés disponible, ainsi que le pourcentage de données manquantes ou aberrantes.

	Code station	Date du premier relevé	Date du dernier relevé	Nombre de relevés	% données manquantes ou aberrantes
Ajaccio	300	1981 - 206	2010 - 244	85 248	67%
Boucau-Bayonne	94	1967 - 142	2010 - 244	305 184	20%
Boulogne-sur-mer	111	1941 - 324	2010 - 243	198 912	67%
Brest	3	1846 - 4	2010 - 243	1 318 896	9%
Calais	55	1941 - 123	2010 - 243	285 216	53%
Cherbourg	13	1943 - 89	2010 - 244	320 040	46%
Concarneau	160	1971 - 201	2010 - 243	104 304	70%
Dieppe	24	1953 - 365	2010 - 244	302 232	40%
Dunkerque	2	1956 - 158	2010 - 243	368 016	23%
Arcachon-Eyrac	190	1967 - 331	2010 - 244	101 808	73%
Le Havre	4	1938 - 4	2010 - 244	357 912	44%
Le Conquet	152	1970 - 355	2010 - 243	334 824	4%
Le Crouesty	185	1996 - 7	2010 - 244	79 560	38%
La Rochelle	34	1941 - 138	2010 - 244	234 648	62%
Marseille	524	1985 - 1	2010 - 50	114 096	48%
Monaco	22	1960 - 1	2010 - 245	155 256	65%
Nice	339	1981 - 184	2010 - 244	116 736	54%
Les Sables- d'Olonne	62	1965 - 186	2010 - 244	181 848	54%
Port-Bloc	15	1959 - 108	2010 - 133	196 896	56%
Port-Vendres	75	1981 - 362	2010 - 242	34 056	86%
Roscoff	54	1973 - 96	2010 - 243	309 216	6%
Sete	250	1992 - 7	2010 - 243	129 552	21%
Saint-Jean-de-Luz	95	1942 - 324	2010 - 243	336 960	44%
Saint-Malo	410	1986 - 99	2010 - 243	143 472	33%
Saint-Nazaire	37	1970 - 1	2010 - 244	182 832	49%
Toulon	68	1961 - 1	2010 - 244	224 208	49%

Tableau 1 : Taille de l'historique et pourcentage de données manquantes pour les 26 stations disponibles

La station de Brest se démarque des autres : il s'agit de la station dont l'historique est le plus étendu (1846-2010) et le plus fourni (9% de données manquantes).

On constate une variabilité non seulement dans l'étendue de l'historique (date de début), ainsi que dans la quantité de données réellement exploitables : deux stations de début d'historique

identique peuvent avoir des historiques de qualités différentes, en termes de nombre de données non manquantes ou aberrantes.

La quasi-totalité des historiques débute après 1940 ; plus de la moitié (62%) ne débute qu'après 1960.

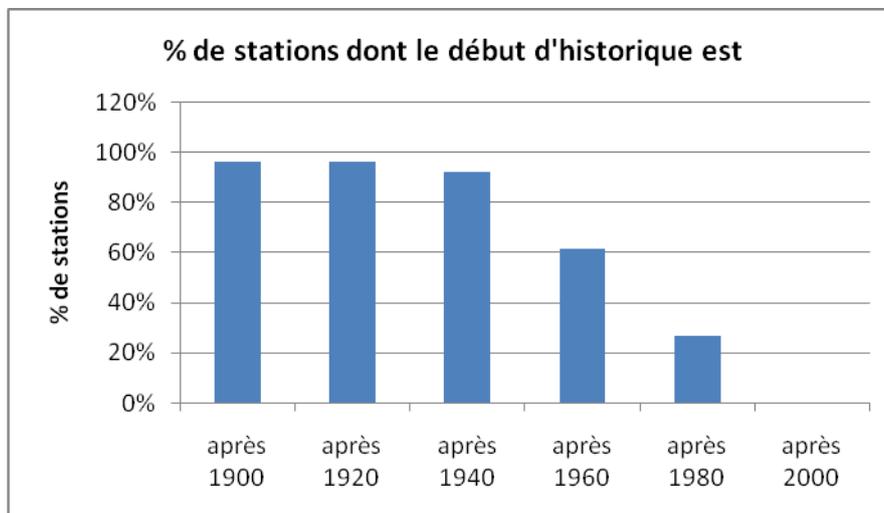


Figure 2 : Qualité des historiques disponibles : répartition des dates de début de relevé

Seules 6 stations possèdent un historique complet à plus de 75%. 42% d'entre elles ont au moins la moitié de données manquantes.

Nombre de stations dont le pourcentage de données manquantes est compris entre	
0-25%	6
25-50%	9
50-75%	10
75-100%	1

Tableau 2 : Qualité des historiques disponibles : répartition des % de données

La figure ci-dessous représente l'évolution du % de données manquantes en fonction de la date de début d'historique considéré. Ce pourcentage est inférieur à 50% à partir de 1965 seulement.

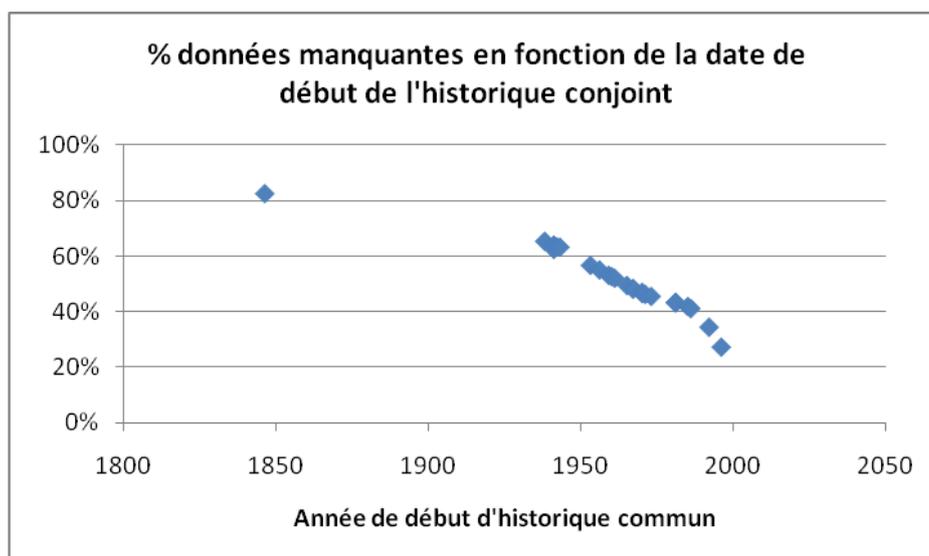


Figure 3 : Qualité des historiques disponibles : % de données manquantes en fonction de la date de début d'historique considérée

Seuls 785 créneaux comportent des mesures communes aux 26 stations, parmi les 1.5 millions de créneaux.

#### D. Analyse statistique après sélection des hauteurs maximales mensuelles

Après analyse des données, on se rend compte qu'une précision horaire n'est pas nécessaire. Nous choisissons d'utiliser comme indicateur la hauteur maximale mensuelle observée sur chaque station, et ce pour tous les mois de l'historique.

Ce choix peut permettre de diminuer la quantité de données manquantes ; de plus, cette précision est cohérente avec celle des données HYDRO des cours d'eau côtiers (hauteurs mensuelles).

Ainsi, si on isole les hauteurs maximales mensuelles pour chaque station, les statistiques obtenues concernant la disponibilité des données sont :

	Année début historique	Mois début	Année fin	Mois fin	nb données théorique	nb données	nb données manquantes (aberrantes)	% données manquantes
AJACC	1981	7	2010	6	348	121	227	65%
BOUCA	1967	5	2010	6	518	431	87	17%
BOULO	1941	11	2010	6	824	295	529	64%
BREST	1846	1	2010	6	1974	1824	150	8%
CALAI	1941	5	2010	6	830	422	408	49%
CHERB	1943	3	2010	6	808	451	357	44%
CONCA	1971	7	2010	6	468	146	322	69%
DIEPP	1953	12	2010	6	679	441	238	35%
DUNKE	1956	6	2010	6	649	531	118	18%

EYRAC	1967	11	2010	6	512	150	362	71%
HAVRE	1938	1	2010	6	870	512	358	41%
LCONQ	1970	12	2010	6	475	466	9	2%
LCROU	1996	3	2010	6	172	116	56	33%
LROCH	1941	5	2010	6	830	336	494	60%
MARSE	1985	1	2010	6	306	162	144	47%
MONAC	1960	1	2010	6	606	216	390	64%
NICE1	1981	7	2010	6	348	163	185	53%
OLONE	1965	7	2010	6	540	257	283	52%
PBLOC	1959	4	2010	6	615	289	326	53%
PVEND	1981	12	2010	6	343	48	295	86%
RSCOF	1973	4	2010	6	447	435	12	3%
SETE1	1992	1	2010	6	222	184	38	17%
SJLUZ	1942	11	2010	6	812	477	335	41%
SMALO	1986	4	2010	6	291	199	92	32%
SNAZA	1970	1	2010	6	486	260	226	47%
TOULO	1961	1	2010	6	594	323	271	46%

Tableau 3 : Taille de l'historique et pourcentage de données manquantes pour les 26 stations disponibles après sélection des hauteurs maximales mensuelles

La proportion de données manquantes observée pour chaque station baisse de manière peu significative. Ceci s'explique par le fait que les données manquantes correspondent à une absence de mesures pouvant s'étendre sur des mois, voire des années, et non pas des pannes de quelques heures ou quelques jours, difficulté que la prise en compte de données mensuelles permettrait de contourner.

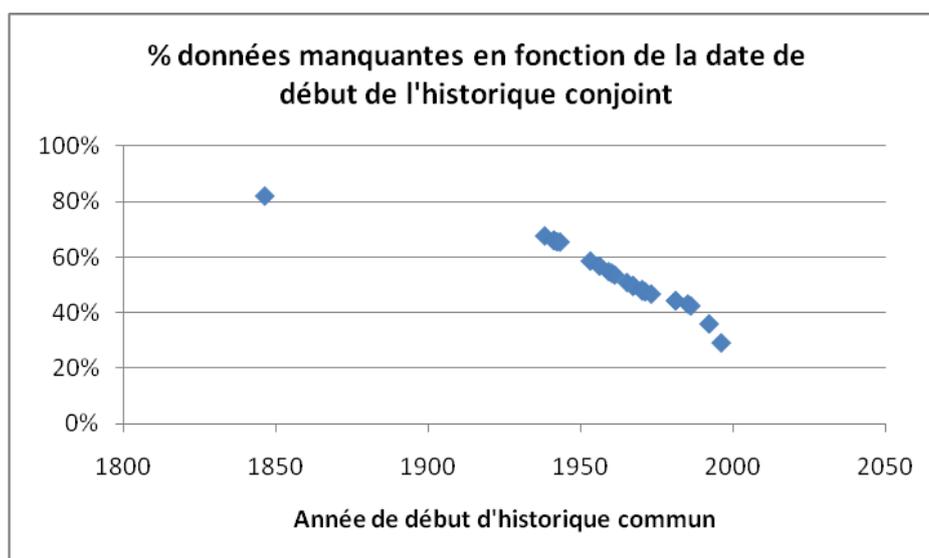


Figure 4 : Qualité des historiques disponibles : % de données manquantes en fonction de la date de début d'historique considérée

Le pourcentage de données manquantes en fonction de la date de début d'historique considéré ne varie pas de façon notable lorsqu'on passe des données journalières au maximum mensuel.

En effet l'évolution est similaire, et le pourcentage est inférieur à 50% à partir de 1965 seulement, de la même façon que pour les données journalières.

### *E. Etude de la corrélation entre les stations SHOM*

La quantité de données manquantes étant importante, nous étudions les corrélations entre stations (deux à deux) : en cas de très forte corrélation, un couple peut être remplacé par une station unique, synthétisant l'information contenue dans les deux historiques. Dans ce cas, la station conservée sera celle dont le nombre de hauteurs extrêmes réellement observées est le plus important (en comparaison au nombre de hauteurs extrêmes reconstruites).

Avant d'étudier la corrélation entre  $x_i$  et  $y_i$ , nous les centrons sur la moyenne ; c'est-à-dire que nous considérons :  $x_i = x_i - \bar{x}$  et  $y_i = y_i - \bar{y}$

Les résultats du calcul de ce coefficient pour les stations situées sur la côte méditerranéenne et sur la côte atlantique sont donnés dans les tableaux ci-dessous.

	<b>AJACC</b>	<b>MARSE</b>	<b>MONAC</b>	<b>NICE1</b>	<b>PVEND</b>	<b>SETE1</b>	<b>TOULO</b>
<b>AJACC</b>	1	0,65705816	0,74057647	0,76715659	0,6820574	0,65072987	0,71052116
<b>MARSE</b>	0,65705816	1	0,69566208	0,6956258	0,70921211	0,62721921	0,6796533
<b>MONAC</b>	0,74057647	0,69566208	1	0,9147552	0,72199124	0,7260338	0,88331225
<b>NICE1</b>	0,76715659	0,6956258	0,9147552	1	0,84144477	0,73246698	0,86781014
<b>PVEND</b>	0,6820574	0,70921211	0,72199124	0,84144477	1	0,94933661	0,78756066
<b>SETE1</b>	0,65072987	0,62721921	0,7260338	0,73246698	0,94933661	1	0,81091403
<b>TOULO</b>	0,71052116	0,6796533	0,88331225	0,86781014	0,78756066	0,81091403	1

*Tableau 4 : Coefficients de corrélation « classique »  
entre les différentes stations SHOM situées sur la côte méditerranéenne*

	BOUCA	BOULO	BREST	CALAI	CHERB	CONCA	DIEPP	DUNKE	EYRAC	HAVRE	LCONQ	LCROU	LROCH	OLONE	PBLOC	RSCOF	SJLUZ	SMALO	SNAZA
BOUCA	1	0,5610	0,6948	0,2524	0,7328	0,7460	0,4231	0,2189	0,2700	0,5838	0,7440	0,2971	0,5827	0,8348	0,2992	0,7247	0,6835	0,0084	0,4458
BOULO	0,5610	1	0,6371	0,5836	0,6627	0,4201	0,7250	0,3750	0,6297	0,6905	0,5651	0,6537	0,5641	0,5438	0,3959	0,5987	0,5613	-0,0329	0,4372
BREST	0,6948	0,6371	1	0,3765	0,6809	0,8382	0,6484	0,3507	0,3431	0,6761	0,9018	0,3962	0,7602	0,8683	0,6110	0,8707	0,7611	0,0480	0,5951
CALAI	0,2524	0,5836	0,3765	1	0,3522	0,3963	0,2943	0,5364	0,6187	0,4341	0,3587	0,1270	0,4664	0,3689	0,2517	0,4007	0,2818	0,0066	0,4145
CHERB	0,7328	0,6627	0,6809	0,3522	1	0,8433	0,3713	0,2233	0,3560	0,5940	0,8264	0,4118	0,7100	0,8508	0,6762	0,7909	0,6989	0,0451	0,5315
CONCA	0,7460	0,4201	0,8382	0,3963	0,8433	1	0,4317	0,3891	0,3102	0,6774	0,7811	0,4375	0,7762	0,8717	0,5857	0,7678	0,7086	-0,0409	0,7448
DIEPP	0,4231	0,7250	0,6484	0,2943	0,3713	0,4317	1	0,3213	0,4178	0,4771	0,5213	0,6300	0,5478	0,3336	0,4215	0,5168	0,4964	0,0333	0,6212
DUNKE	0,2189	0,3750	0,3507	0,5364	0,2233	0,3891	0,3213	1	0,2191	0,3867	0,2947	0,2392	0,5033	0,5607	0,3343	0,2451	0,2550	0,1117	0,2860
EYRAC	0,2700	0,6297	0,3431	0,6187	0,3560	0,3102	0,4178	0,2191	1	0,3463	0,3479	0,2652	0,1374	0,3299	0,2949	0,3372	0,5442	0,0019	0,2341
HAVRE	0,5838	0,6905	0,6761	0,4341	0,5940	0,6774	0,4771	0,3867	0,3463	1	0,6210	0,3322	0,6629	0,6857	0,5271	0,6765	0,5073	0,0732	0,4998
LCONQ	0,7440	0,5651	0,9018	0,3587	0,8264	0,7811	0,5213	0,2947	0,3479	0,6210	1	0,3885	0,7711	0,8744	0,5362	0,8048	0,7606	0,0046	0,5166
LCROU	0,2971	0,6537	0,3962	0,1270	0,4118	0,4375	0,6300	0,2392	0,2652	0,3322	0,3885	1	0,4163	0,4126	0,6756	0,3490	0,7998	-0,0190	0,9346
LROCH	0,5827	0,5641	0,7602	0,4664	0,7100	0,7762	0,5478	0,5033	0,1374	0,6629	0,7711	0,4163	1	0,8444	0,5930	0,7198	0,6390	0,0177	0,7981
OLONE	0,8348	0,5438	0,8683	0,3689	0,8508	0,8717	0,3336	0,5607	0,3299	0,6857	0,8744	0,4126	0,8444	1	0,8238	0,8474	0,7150	0,0700	0,9172
PBLOC	0,2992	0,3959	0,6110	0,2517	0,6762	0,5857	0,4215	0,3343	0,2949	0,5271	0,5362	0,6756	0,5930	0,8238	1	0,5863	0,5214	-0,1263	0,6020
RSCOF	0,7247	0,5987	0,8707	0,4007	0,7909	0,7678	0,5168	0,2451	0,3372	0,6765	0,8048	0,3490	0,7198	0,8474	0,5863	1	0,7537	0,0039	0,4845
SJLUZ	0,6835	0,5613	0,7611	0,2818	0,6989	0,7086	0,4964	0,2550	0,5442	0,5073	0,7606	0,7998	0,6390	0,7150	0,5214	0,7537	1	-0,0840	0,4949
SMALO	0,0084	-0,0329	0,0480	0,0066	0,0451	-0,0409	0,0333	0,1117	0,0019	0,0732	0,0046	-0,0190	0,0177	0,0700	-0,1263	0,0039	-0,0840	1	-0,2230
SNAZA	0,4458	0,4372	0,5951	0,4145	0,5315	0,7448	0,6212	0,2860	0,2341	0,4998	0,5166	0,9346	0,7981	0,9172	0,6020	0,4845	0,4949	-0,2230	1

$x_i$  et  $y_i$  sont centrés sur la moyenne.

Ainsi nous cherchons pour chaque station présentant un grand nombre de données manquantes si elle est fortement corrélée avec l'une de ses voisines afin de pouvoir l'éliminer. Ces coefficients de corrélation permettent également de déterminer s'il est possible de compléter les données manquantes d'une station à l'aide des stations situées le long de la même côte. Plus le coefficient de corrélation est proche de 1, plus la reconstruction d'une station par une autre est de bonne qualité.

Les coefficients de corrélation entre les stations de la côte méditerranéenne sont :

	<b>AJACC</b>	<b>MARSE</b>	<b>MONAC</b>	<b>NICE1</b>	<b>PVEND</b>	<b>SETE1</b>	<b>TOULO</b>
<b>AJACC</b>	1	0,59668589	0,72095223	0,72071457	0,49587336	0,57841246	0,71035391
<b>MARSE</b>	0,47580933	1	0,65975307	0,6616013	0,30069268	0,56400315	0,64214728
<b>MONAC</b>	0,52249063	0,4312983	1	0,69143669	0,42737247	0,42325839	0,7543309
<b>NICE1</b>	0,64668587	0,61510852	0,8428273	1	0,49290639	0,63705705	0,85879048
<b>PVEND</b>	0,59826313	0,4579025	0,70802971	0,67069958	1	0,61788857	0,6916237
<b>SETE1</b>	0,10080025	0,11357344	0,12816507	0,13718164	0,08267543	1	0,19777242
<b>TOULO</b>	0,47561954	0,40271385	0,70151825	0,61199124	0,3903884	0,56877322	1

Tableau 5 : Coefficients de corrélation entre les différentes stations SHOM situées sur la côte méditerranéenne

Le calcul des coefficients de corrélation entre les stations situées sur les côtes atlantiques et de la manche donne :

	BOUCA	BOULO	BREST	CALAI	CHERB	CONCA	DIEPP	DUNKE	EYRAC	HAVRE	LCONQ	LCROU	LROCH	OLONE	PBLOC	RSCOF	SJLUZ	SMALO	SNAZA
<b>BOUCA</b>	1	0,3919	0,6948	0,2299	0,6046	0,3559	0,3479	0,2007	0,1420	0,4827	0,6171	0,1256	0,4673	0,5146	0,2169	0,5911	0,6447	0,0044	0,2932
<b>BOULO</b>	0,5108	1	0,6371	0,4786	0,6374	0,2180	0,5326	0,3186	0,3424	0,6691	0,5408	0,3089	0,3967	0,3589	0,2344	0,5760	0,4997	-0,0172	0,3458
<b>BREST</b>	0,3296	0,2472	1	0,1748	0,3298	0,2189	0,3080	0,1822	0,1021	0,3429	0,4383	0,0938	0,3006	0,3062	0,2294	0,4095	0,3805	0,0145	0,2242
<b>CALAI</b>	0,2296	0,3410	0,3765	1	0,3143	0,1587	0,2460	0,3804	0,2610	0,3967	0,3272	0,0429	0,2738	0,2911	0,1363	0,2732	0,2632	0,0048	0,2251
<b>CHERB</b>	0,4772	0,3795	0,6809	0,2056	1	0,3161	0,3202	0,2087	0,1437	0,5846	0,5688	0,1413	0,3617	0,4361	0,5351	0,5448	0,4472	0,0201	0,2770
<b>CONCA</b>	0,6420	0,3452	0,8382	0,3541	0,7321	1	0,2323	0,3666	0,2263	0,6724	0,7713	0,3038	0,7334	0,6954	0,5150	0,7416	0,5851	-0,0327	0,5980
<b>DIEPP</b>	0,3383	0,4493	0,6484	0,2277	0,2859	0,1136	1	0,2607	0,1776	0,3760	0,4107	0,1518	0,2699	0,1630	0,2235	0,3633	0,4259	0,0146	0,4131
<b>DUNKE</b>	0,1960	0,3178	0,3507	0,4857	0,1993	0,1250	0,2937	1	0,0771	0,3585	0,2683	0,0662	0,2505	0,2301	0,1737	0,2195	0,2375	0,0374	0,1420
<b>EYRAC</b>	0,2685	0,2337	0,3431	0,2959	0,3383	0,2873	0,1663	0,2149	1	0,3318	0,3418	0,2421	0,1298	0,3062	0,2851	0,3204	0,2536	0,0017	0,2254
<b>HAVRE</b>	0,5002	0,4976	0,6670	0,3396	0,5405	0,3154	0,3490	0,3258	0,1599	1	0,5704	0,1237	0,4303	0,4583	0,2941	0,5957	0,4367	0,0399	0,3425
<b>LCONQ</b>	0,6842	0,4506	0,9018	0,3028	0,7888	0,4160	0,3836	0,2641	0,1928	0,6075	1	0,1864	0,5567	0,6111	0,3269	0,7715	0,6904	0,0027	0,3899
<b>LCROU</b>	0,2943	0,2088	0,3962	0,1243	0,4118	0,4324	0,1190	0,2383	0,1021	0,3322	0,3885	1	0,4156	0,4126	0,3935	0,3484	0,2679	-0,0190	0,2600
<b>LROCH</b>	0,4805	0,3703	0,7547	0,4368	0,5325	0,4786	0,4418	0,4796	0,0785	0,5251	0,6762	0,2257	1	0,5400	0,4849	0,5584	0,5694	0,0108	0,6195
<b>OLONE</b>	0,7745	0,3886	0,8683	0,3140	0,8407	0,6262	0,1613	0,4809	0,2406	0,6774	0,8580	0,2819	0,6942	1	0,5733	0,8333	0,5987	0,0603	0,4787
<b>PBLOC</b>	0,2387	0,2264	0,6110	0,2181	0,4610	0,4068	0,3786	0,2970	0,1984	0,4566	0,4289	0,3176	0,5040	0,4331	1	0,3709	0,4399	-0,0650	0,4464
<b>RSCOF</b>	0,6683	0,5036	0,8707	0,3353	0,7801	0,3975	0,3840	0,2169	0,1827	0,6679	0,7949	0,1621	0,5292	0,5653	0,3741	1	0,6923	0,0023	0,3824
<b>SJLUZ</b>	0,5902	0,3985	0,7594	0,2288	0,6184	0,2699	0,4028	0,2107	0,2608	0,4403	0,6553	0,2975	0,4417	0,4188	0,3238	0,6211	1	-0,0407	0,3461
<b>SMALO</b>	0,0078	-0,0184	0,0480	0,0048	0,0451	-0,0273	0,0200	0,0892	0,0012	0,0732	0,0045	-0,0128	0,0138	0,0696	-0,0827	0,0039	-0,0682	1	-0,0916
<b>SNAZA</b>	0,4331	0,4151	0,5951	0,2310	0,5078	0,2429	0,3635	0,2827	0,0815	0,4877	0,5072	0,2308	0,3953	0,2623	0,2469	0,4703	0,4880	-0,0660	1

Tableau 6 : Coefficients de corrélation entre les différentes stations SHOM situées sur la côte Atlantique et la Manche

Les coefficients obtenus permettent de hiérarchiser la corrélation entre chaque station. Ainsi, lors du choix des stations, ce classement permettra d'éliminer les stations comportant une grande proportion de données manquantes et étant fortement corrélées à des stations comportant moins de données manquantes. Cependant ce type de station sera utilisé pour la reconstruction d'autres stations.

### F. Détection d'évènements extrêmes

Afin de procéder à la sélection des stations SHOM pour l'étude, il est nécessaire de détecter les stations contenant des valeurs extrêmes de hauteurs d'eau, afin de les prendre en compte par la suite. En effet, l'objectif du contrat est d'évaluer la probabilité d'évènements extrêmes ; il est donc particulièrement intéressant de sélectionner des stations présentant le plus possible de phénomènes extrêmes « réels », et non pas reconstruits virtuellement.

Dans un premier temps, la loi de probabilité des hauteurs maximales mensuelles est calculée pour chaque station. Des histogrammes de répartition des hauteurs sont donc obtenus. Par exemple, pour la station SHOM d'Ajaccio, on obtient :

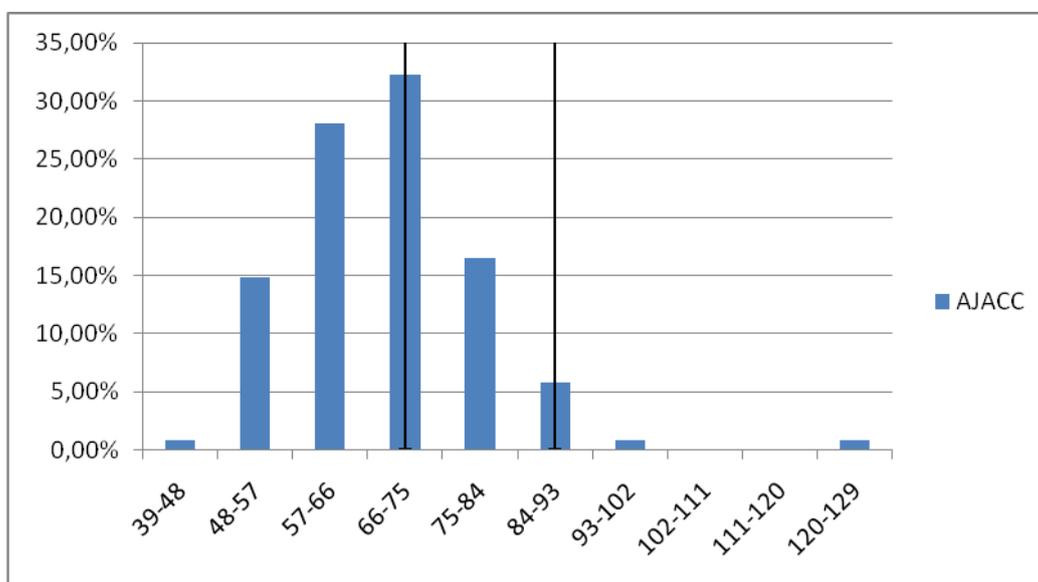


Figure 5 : Répartition des hauteurs d'eau maximales mensuelles pour la station d'Ajaccio

Pour cette station, on peut constater un étalement de la répartition de la loi vers la droite, il semble donc que des valeurs extrêmes aient été mesurées.

De plus, si l'on trace les valeurs de la médiane et du quantile à 99% (barres noires verticales sur la figure), nous pouvons visualiser si la dernière classe est très éloignée de la classe incluant le quantile à 99%.

La différence entre la valeur maximale mesurée et le quantile à 99 % peut être quantifiée sous forme de cette différence relative :

$$diff\_relative = \frac{valeur\_maxi - quantile\_99}{valeur\_maxi - valeur\_mini}$$

Cette valeur se présente sous la forme d'un pourcentage correspondant à « l'étalement » de la distribution des valeurs de hauteur vers la droite.

Station	Différence relative
EYRAC	70,85%
AJACC	37,44%
OLONE	30,81%
LROCH	29,26%
CONCA	13,67%
PBLOC	15,36%
MONAC	13,79%
SETE1	11,41%
DIEPP	10,58%
TOULO	9,84%
NICE1	8,32%
SNAZA	8,46%
RSCOF	7,41%
HAVRE	6,54%
SMALO	6,84%
LCROU	1,66%
DUNKE	5,65%
LCONQ	5,59%
BREST	5,37%
SJLUZ	4,65%
BOUCA	3,19%
MARSE	3,37%
CALAI	2,88%
CHERB	2,71%
BOULO	1,44%
PVEND	0,00%

*Tableau 7 : Différence relative entre la hauteur maximale et le quantile à 99% par rapport à l'étendue de l'échantillon*

Nous dirons qu'il y a évènement extrême lorsque la différence relative est supérieure à 10%. Les stations correspondant à ce critère sont :

- Arcachon – Eyrac ;
- Ajaccio ;
- Les Sables d'Olonne ;
- La Rochelle ;
- Concarneau ;
- Port-Bloc ;
- Monaco ;
- Sète ;
- Dieppe.

Les lois de probabilités de ces stations sont présentées sur les figures ci-dessous :

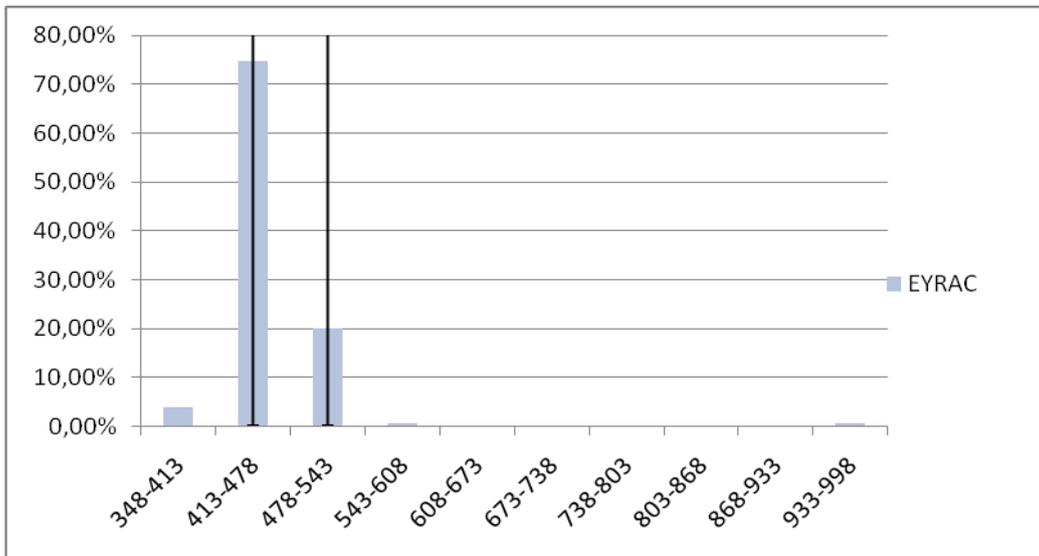


Figure 6 : Loi de probabilité de la station Arcachon-Eyrac (médiane et quantile à 99%)

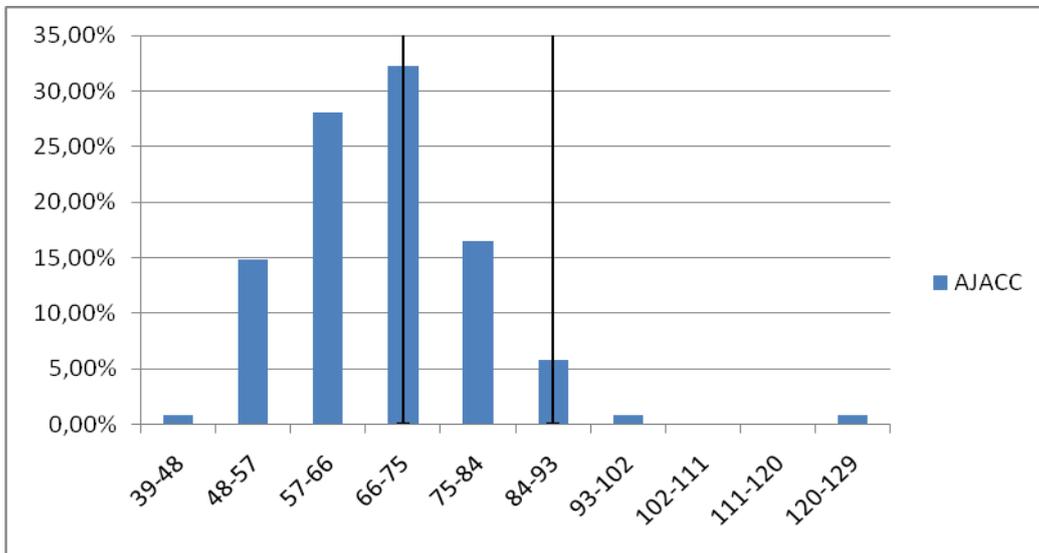


Figure 7 : Loi de probabilité de la station Ajaccio (médiane et quantile à 99%)

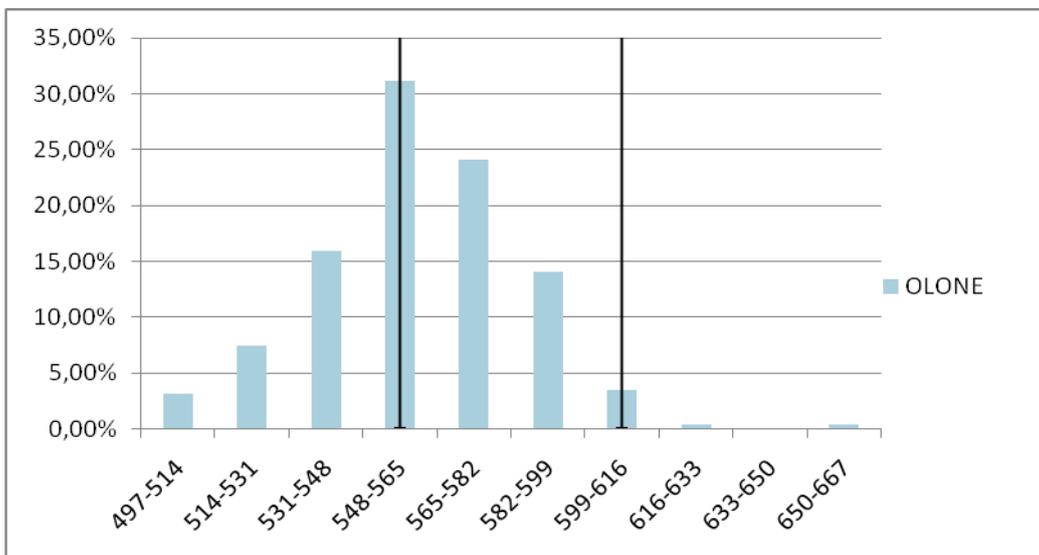


Figure 8 : Loi de probabilité de la station Les Sables d'Olonne (médiane et quantile à 99%)

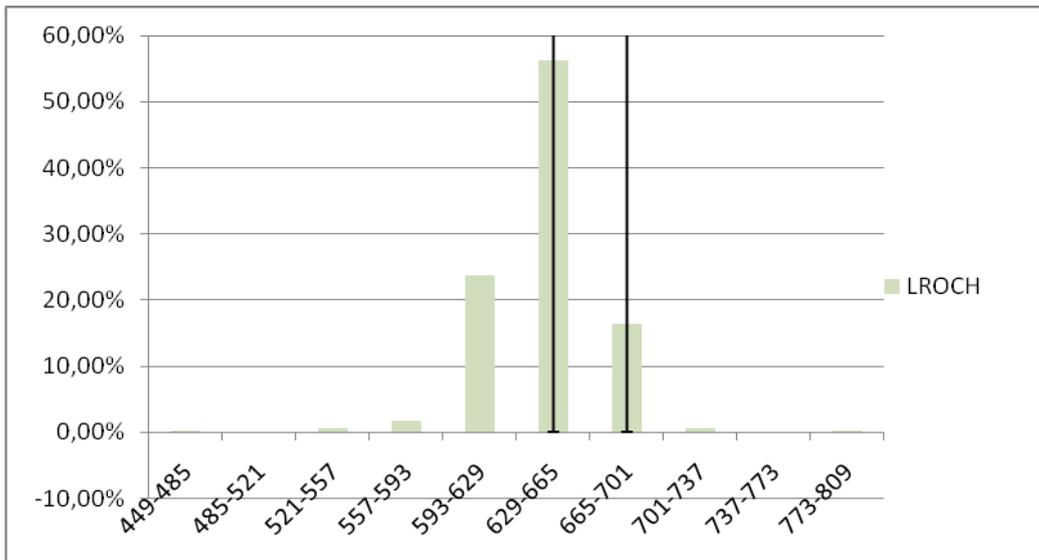


Figure 9 : Loi de probabilité de la station la Rochelle (médiane et quantile à 99%)

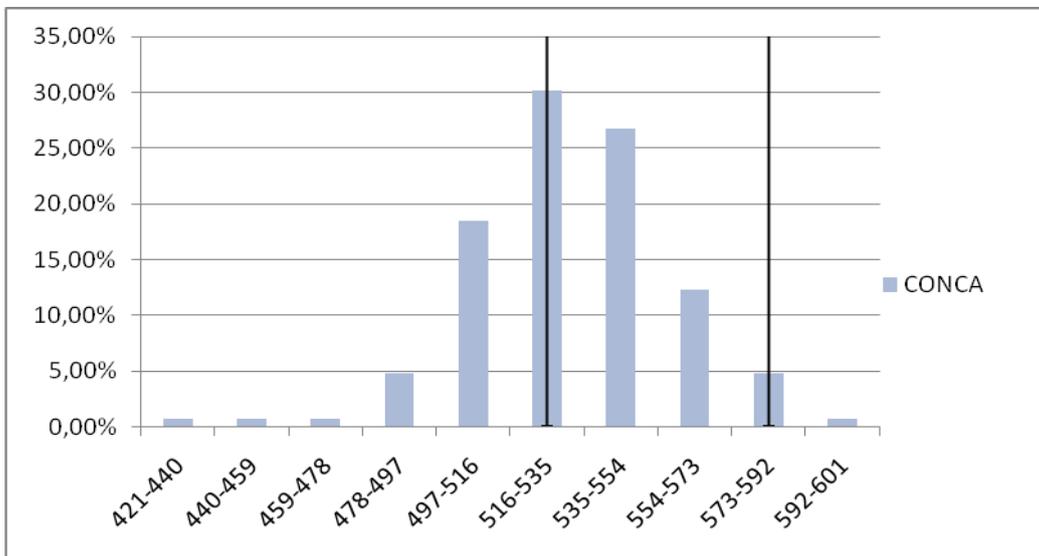


Figure 10 : Loi de probabilité de la station Concarneau (médiane et quantile à 99%)

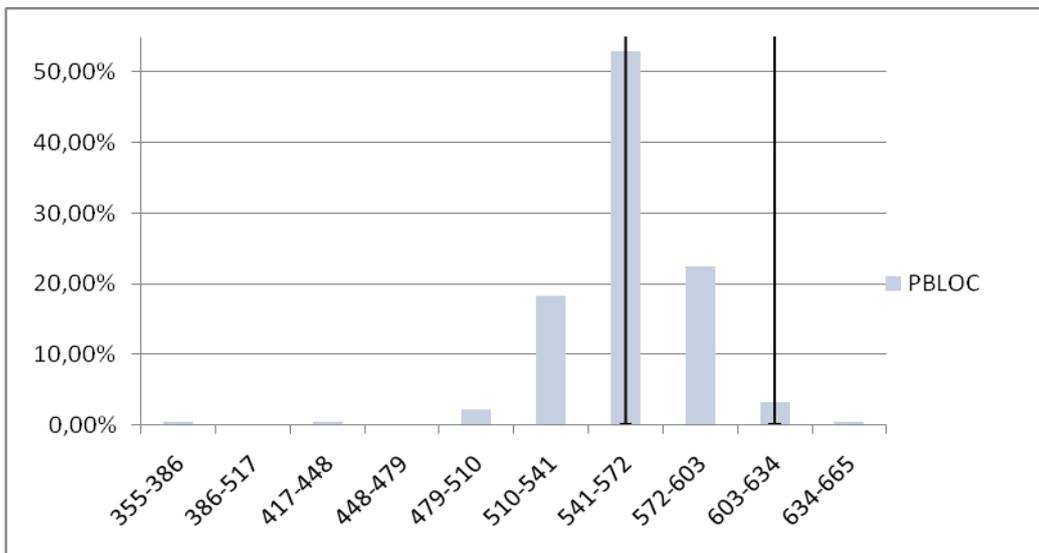


Figure 11 : Loi de probabilité de la station Port Bloc (médiane et quantile à 99%)

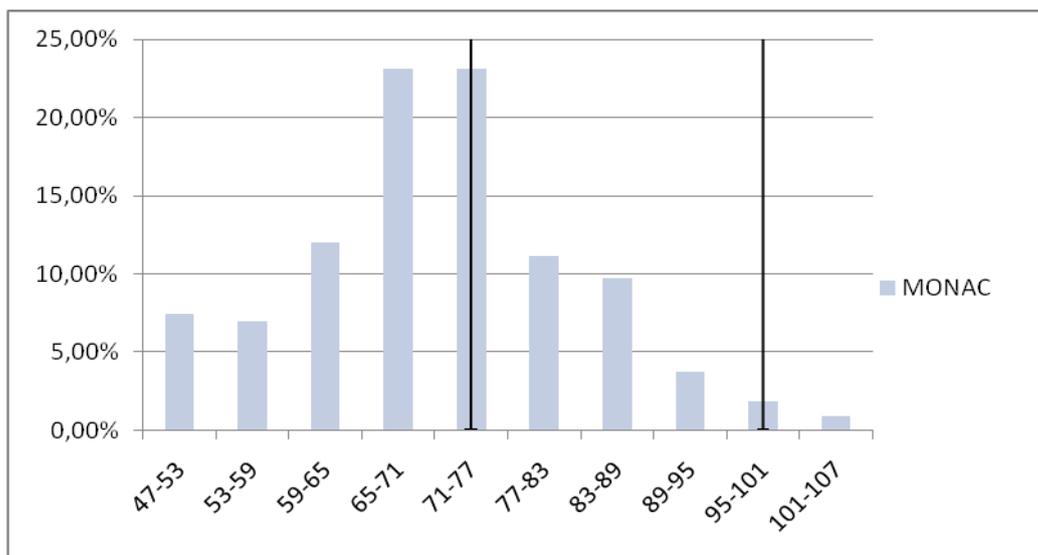


Figure 12 : Loi de probabilité de la station Monaco (médiane et quantile à 99%)

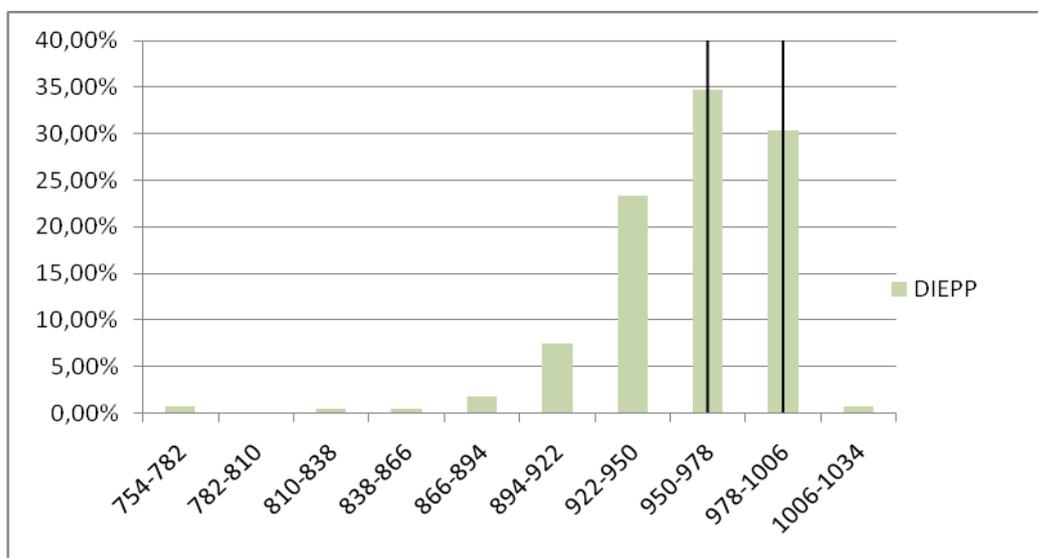


Figure 13 : Loi de probabilité de la station Dieppe (médiane et quantile à 99%)

Les représentations graphiques ci-dessus permettent donc de confirmer l'observation d'évènements considérés comme extrêmes.

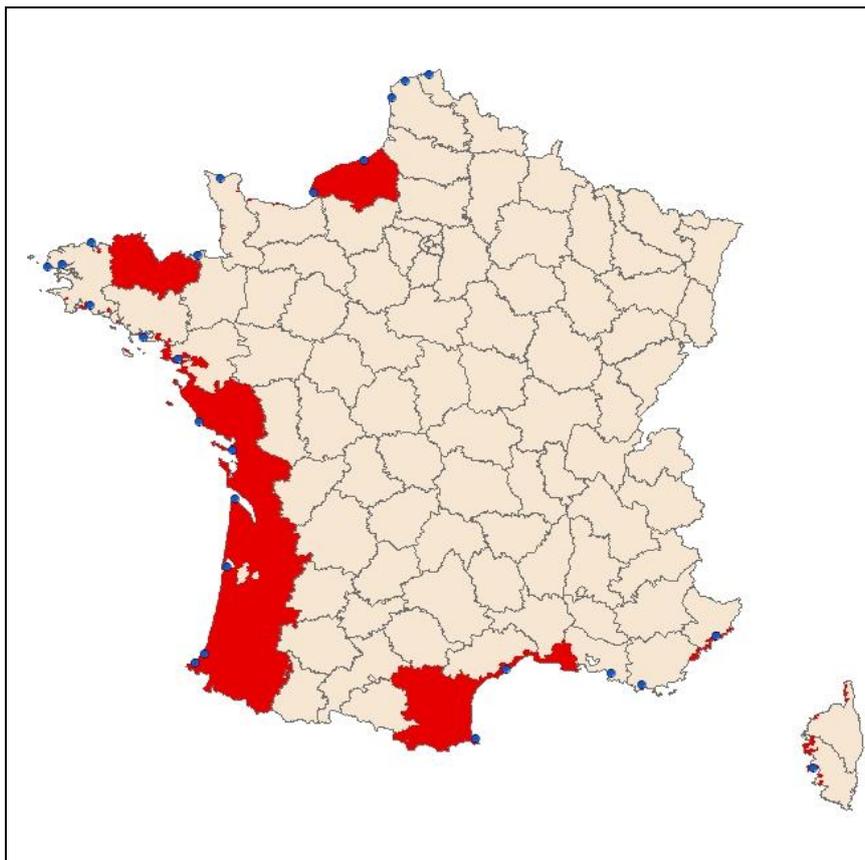
Après avoir défini ces évènements, il sera nécessaire d'en tenir compte lors de la sélection des stations à conserver dans l'étude menée.

### *G. Prise en compte de la houle : croisement des données avec celles de l'historique CMV de la CCR*

Une question se pose concernant les données : prennent-elles en compte l'effet de la houle ? En effet, la présence de houle peut augmenter considérablement l'amplitude des vagues et provoquer des inondations, qui n'auraient pas lieu pour une hauteur d'eau moyenne similaire sans houle. Il est donc important de savoir si les données SHOM prennent en compte ce phénomène ou non. Si tel n'est pas le cas, les données relatives à la houle pourront être demandées auprès du CETMEF par la CCR.

Pour ce faire, nous croisons les données de hauteurs SHOM aux relevés d'évènements CMV (Chocs Mécaniques dus aux Vagues) de la CCR : il s'agit de l'historique des évènements CMV, dont font parties les inondations qui nous intéressent. Les données contiennent la date de survenue de l'évènement, ainsi que la commune où a eu lieu la catastrophe.

En France métropolitaine, 7 tempêtes ont frappé les côtes, entre 1997 et 2010 (Xynthia). Nous avons utilisé Arcgis pour représenter la localisation des communes touchées par au moins une tempête, et les superposer aux marégraphes du SHOM :



*Figure 14 : Localisation des tempêtes ayant provoqué des inondations et stations de mesure SHOM*

## 1. Première étude

Afin d'étudier si la houle est prise en compte ou non, nous isolons tout d'abord dans l'historique des événements CMV, ceux qui ont été enregistrés dans une commune disposant d'un marégraphe du SHOM.

Nous obtenons le tableau ci-dessous :

Libellé	Evt. Géo.	Début Rec.	Fin Rec.
Nice (06)	G_200011_ICB_NICE	5 novembre 2000	6 novembre 2000
La Rochelle / Rochelle 1er Canton (17)	G_199912_ICB_LOTHAR_ET_MARTIN	25 décembre 1999	29 décembre 1999
La Rochelle (17)	G_201002_ICB_XYNTHIA	27 février 2010	1 mars 2010
Ajaccio (2A)	G_199912_ICB_LOTHAR_ET_MARTIN	28 décembre 1999	28 décembre 1999
Arcachon (33)	G_199912_ICB_LOTHAR_ET_MARTIN	25 décembre 1999	29 décembre 1999
Arcachon (33)	G_200901_ICB_TEMPETE_SUDOUEST	24 janvier 2009	27 janvier 2009
Arcachon (33)	G_201002_ICB_XYNTHIA	27 février 2010	28 février 2010
Sète (34)	G_199712_ICB_LANGUEDOC	16 décembre 1997	19 décembre 1997
Saint-Nazaire (44)	G_201002_ICB_XYNTHIA	28 février 2010	28 février 2010
Bayonne (64)	G_199912_ICB_LOTHAR_ET_MARTIN	25 décembre 1999	29 décembre 1999
Bayonne (64)	G_200901_ICB_TEMPETE_SUDOUEST	24 janvier 2009	27 janvier 2009
Boucau (64)	G_199912_ICB_LOTHAR_ET_MARTIN	25 décembre 1999	29 décembre 1999
Boucau (64)	G_200901_ICB_TEMPETE_SUDOUEST	24 janvier 2009	27 janvier 2009
Saint-Jean-de-Luz (64)	G_199912_ICB_LOTHAR_ET_MARTIN	25 décembre 1999	29 décembre 1999
Saint-Jean-de-Luz (64)	G_200901_ICB_TEMPETE_SUDOUEST	24 janvier 2009	27 janvier 2009
Port-Vendres (66)	G_199712_ICB_LANGUEDOC	16 décembre 1997	19 décembre 1997
Port-Vendres (66)	G_200312_ICB_ARLES	3 décembre 2003	4 décembre 2003
Port-Vendres (66)	G_200901_ICB_TEMPETE_SUDOUEST	24 janvier 2009	27 janvier 2009
Saint-Nazaire (66)	G_200901_ICB_TEMPETE_SUDOUEST	24 janvier 2009	27 janvier 2009
Dieppe / Dieppe-Est (76)	G_199912_ICB_LOTHAR_ET_MARTIN	25 décembre 1999	29 décembre 1999
Le Havre / Havre 1er Canton (76)	G_199912_ICB_LOTHAR_ET_MARTIN	25 décembre 1999	29 décembre 1999
Boulogne (85)	G_201002_ICB_XYNTHIA	27 février 2010	1 mars 2010
Les Sables-d'Olonne (85)	G_199912_ICB_LOTHAR_ET_MARTIN	25 décembre 1999	29 décembre 1999
Les Sables-d'Olonne (85)	G_201002_ICB_XYNTHIA	27 février 2010	1 mars 2010

Tableau 8 : Liste des communes ayant connu des événements CMV et disposant de marégraphe SHOM

Nous vérifions ensuite dans l'historique des relevés SHOM si les données sont disponibles pour les dates de tempêtes, et conservons uniquement les communes où c'est le cas. Pour chacune de ces communes, nous stockons les hauteurs d'eau minimale et maximale relevées sur la période de la tempête, et nous comparons ces hauteurs aux hauteurs prévues à ces dates par le SHOM.

Libellé	Début Rec.	Fin Rec.	Hauteur min mesurée	Hauteur max mesurée	Hauteur min prévue	Hauteur max prévue
Arcachon	24 janvier 2009	27 janvier 2009	76	488	74	419
Arcachon	27 février 2010	28 février 2010	46	546	25	456
Boucau	25 décembre 1999	29 décembre 1999	90	461	53	451
Boucau	24 janvier 2009	27 janvier 2009	122	443	104	413
La Rochelle	27 février 2010	1 mars 2010	55	800	34	668
La Rochelle / Rochelle 1er Canton	25 décembre 1999	29 décembre 1999	107	670	68	630
Le Havre / Havre 1er Canton	25 décembre 1999	29 décembre 1999	145	888	86	812
Les Sables-d'Olonne	27 février 2010	1 mars 2010	141	659	12	568
Nice	5 novembre 2000	6 novembre 2000	45	83	46	62
Saint-Nazaire	24 janvier 2009	27 janvier 2009	134	594	125	570
Saint-Nazaire	28 février 2010	28 février 2010	74	722	40	618
Sète	16 décembre 1997	19 décembre 1997	41	129	31	56

Tableau 9 : Liste des communes dont les relevés de hauteurs d'eau sont disponibles aux dates des évènements CMV

Les hauteurs prévues par le SHOM sont construites à partir des coefficients de marée ; elles ne prennent pas en compte les perturbations atmosphériques (et donc la houle) :

- Calcul de la marée dans les principaux ports du monde et pour tous les ports, principaux et secondaires, de France métropolitaine.
- Ces prédictions sont calculées avec une précision de quelques centimètres pour les hauteurs et quelques minutes pour les heures ; la hauteur d'eau réelle peut toutefois s'écarter notablement de la prédiction (jusqu'à plusieurs dizaines de centimètres) en raison notamment des variations du niveau de la mer dues aux perturbations atmosphériques.

(Source : [http://www.shom.fr/fr\\_page/fr\\_serv\\_prediction/ann\\_marees.htm](http://www.shom.fr/fr_page/fr_serv_prediction/ann_marees.htm))

Etudions à présent les écarts relatifs entre les hauteurs et les hauteurs réellement observées. Si ces écarts sont importants dans la majorité des cas, cela tendrait à prouver que les relevés horaires dont nous disposons prennent en compte la houle.

Libellé	Début Rec.	Fin Rec.	Ecart relatif min	Ecart relatif max
Arcachon	24 janvier 2009	27 janvier 2009	2,70%	16,47%
Arcachon	27 février 2010	28 février 2010	84,00%	19,74%
Boucau	25 décembre 1999	29 décembre 1999	69,81%	2,22%
Boucau	24 janvier 2009	27 janvier 2009	17,31%	7,26%
La Rochelle	27 février 2010	1 mars 2010	61,76%	19,76%
La Rochelle / Rochelle 1er Canton	25 décembre 1999	29 décembre 1999	57,35%	6,35%
Le Havre / Havre 1er Canton	25 décembre 1999	29 décembre 1999	68,60%	9,36%
Les Sables-d'Olonne	27 février 2010	1 mars 2010	1075,00%	16,02%
Nice	5 novembre 2000	6 novembre 2000	-2,17%	33,87%
Saint-Nazaire	24 janvier 2009	27 janvier 2009	7,20%	4,21%
Saint-Nazaire	28 février 2010	28 février 2010	85,00%	16,83%
Sète	16 décembre 1997	19 décembre 1997	32,26%	130,36%

*Tableau 10 : Ecart relatif entre hauteurs d'eau minimales et maximales prévues par le SHOM et hauteurs d'eau réellement observées aux dates des événements CMV*

Les écarts relatifs observés varient de -2% à 1075% : on ne peut pas tirer de conclusion quant à la prise en compte ou non de la houle dans les données fournies par le SHOM.

Il est donc nécessaire de compléter cette étude par l'étude suivante : pour chaque station de la liste précédente, il faut sélectionner les dates auxquelles la hauteur d'eau prévue est identique à celle prévue le jour de l'évènement CMV et comparer les hauteurs observées à la date de l'évènement dans cet historique (valeur extrême ou non).

## **2. Seconde approche : étude de la station BOUCAU**

Afin de conclure quant à la prise en compte de la houle dans les relevés de hauteur d'eau, nous nous concentrons sur un exemple de station : la station SHOM de Boucau. Le choix porte sur celle-ci car la commune a connu des tempêtes, et son historique à partir de 1970 ne comporte que 20% de données manquantes.

Les prévisions de hauteurs d'eau faites par le SHOM sont récupérées pour chaque jour à partir de janvier 1970. On retient pour chaque jour la hauteur maximale prévue. L'historique des hauteurs d'eau de la station BOUCAU à partir de 1970 est sélectionné, et pour chaque jour la hauteur maximale mesurée est stockée.

La différence relative entre la hauteur prévue et la hauteur mesurée est calculée pour tous les jours où les deux données sont disponibles. On obtient la répartition suivante :

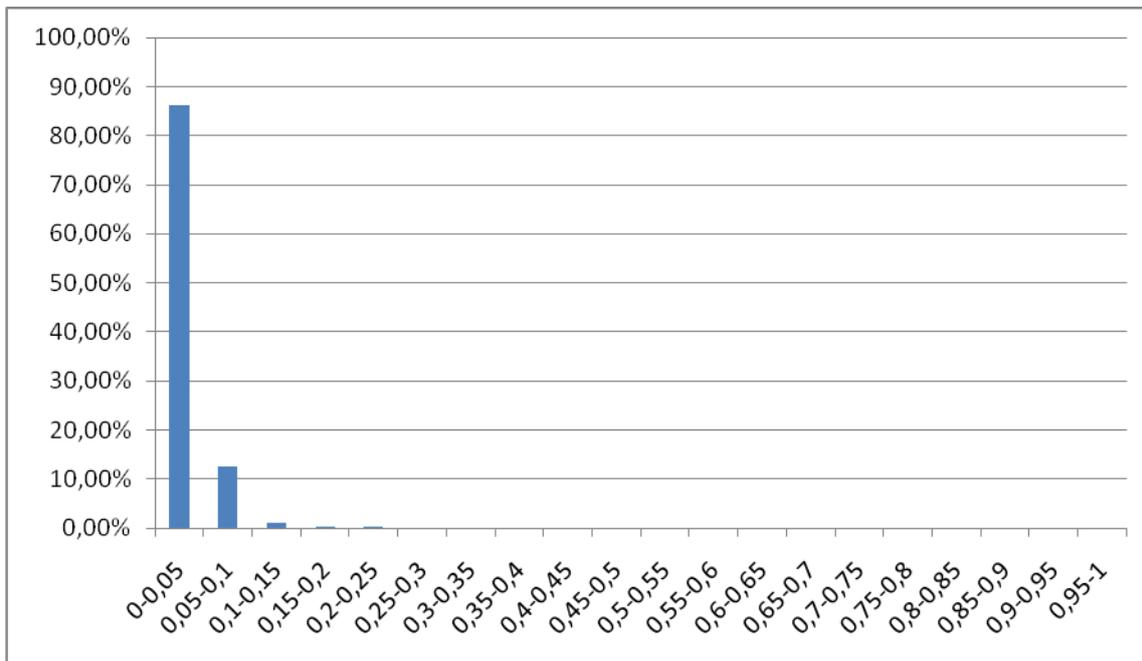


Figure 15 : Différence relative entre hauteur prévue et hauteur mesurée – station de BOUCAU

Plus de 80% des différences relatives calculées sont inférieures à 0.05 ; et la valeur maximale observée est égale à 0.22. De plus, les différences les plus importantes calculées ne correspondent pas à des jours de tempête. Cela signifie que dans la majorité des cas, les hauteurs maximales prévues et observées sont très proches. Etant donné que la prévision des hauteurs ne prend pas en compte l'effet de la houle, on peut en conclure que la houle n'est pas prise en compte lors de la mesure des hauteurs d'eau réalisée par le SHOM.

## H. Corrélation des stations SHOM sur les phénomènes extrêmes

### 1. Corrélation des phénomènes extrêmes

Le but de cette étude est de déterminer s'il existe un lien entre les événements dits « extrêmes » des différentes stations. Ces événements correspondent aux hauteurs d'eau mesurées supérieures à la valeur du quantile à 90% de chaque station.

Une fois les valeurs extrêmes isolées, il reste 332 dates pour lesquelles au moins une station dispose d'une mesure. On calcule ensuite les coefficients de corrélation symétriques entre les stations de la côte Atlantique d'une part et de la côte méditerranéenne d'autre part

Le coefficient de corrélation d'un couple est à croiser avec le nombre d'évènements communs du couple, en effet, un coefficient de 0,99 pour 3 dates communes n'a pas beaucoup de sens.

	AJACC	MARSE	MONAC	NICE1	PVEND	SETE1	TOULO
AJACC	1	0,877	0,826	0,637	1	0,180	0,487
MARSE	0,877	1	0,179	-0,253	1	0,324	-0,088
MONAC	0,826	0,179	1	0,988	1	0,759	0,655
NICE1	0,637	-0,253	0,988	1	0,973	-0,337	0,861
PVEND	1	1	1	0,973	1	-0,832	1
SETE1	0,180	0,324	0,759	-0,337	-0,832	1	-0,403
TOULO	0,487	-0,088	0,655	0,861	1	-0,403	1

Tableau 11 : Coefficients de corrélation des événements extrêmes - Méditerranée

	AJACC	MARSE	MONAC	NICE1	PVEND	SETE1	TOULO
AJACC	11	6	5	6	2	4	7
MARSE	6	16	6	9	2	6	8
MONAC	5	6	20	9	2	4	14
NICE1	6	9	9	17	3	9	10
PVEND	2	2	2	3	5	3	2
SETE1	4	6	4	9	3	16	9
TOULO	7	8	14	10	2	9	33

Tableau 12 : Nombre d'évènements extrêmes simultanés - Méditerranée

Les résultats obtenus pour la côte méditerranéenne confirment dans un premier temps le bien fondé de la suppression de la station Port-Vendres ; en effet, on y observe très peu d'évènements extrêmes (historique court), et elle ne dispose donc que de très peu d'évènements communs avec les autres stations. Elle n'apporte aucune information supplémentaire.

La station d'Ajaccio est bien corrélée à Marseille et Monaco pour respectivement 6 et 5 dates communes, ce qui tend à indiquer que les évènements extrêmes observés à Ajaccio sont similaires à ceux mesurés sur ces stations.

Marseille est uniquement bien corrélée avec Ajaccio, cependant elle dispose de 9 dates communes avec Nice, et 8 avec Toulon, mais les coefficients de corrélation associés à ces couples sont négatifs.

Monaco est bien corrélée avec Nice : 0,988 pour 9 dates, mais faiblement corrélée avec Toulon : 0,655 pour 14 dates.

Pour la station de Toulon, qui est celle disposant du plus grand historique, la seule corrélation pouvant être considérée comme acceptable est avec la station de Nice, avec un coefficient de 0,861 pour 10 dates.

	BOUCA	BOULO	BREST	CALAI	CHERB	CONCA	DIEPP	DUNKE	EYRAC	HAVRE	LCONQ	LCROU	LROCH	OLONE	PBLOC	RSCOF	SJLUZ	SMALO	SNAZA
BOUCA	1	0,401	0,298	-0,192	0,415	0,536	0,168	-0,220	0,698	0,681	0,195	0,783	0,268	0,548	0,655	0,578	0,239	-0,415	0,177
BOULO	0,401	1	0,039	-0,002	0,010	-0,514		0,486	-0,400	0,551	-0,370	-0,091	-0,671	-0,695	0,204	0,764	0,176		0,216
BREST	0,298	0,039	1	0,014	0,474	0,563	0,069	-0,174	0,067	-0,053	0,722	0,475	0,288	0,293	0,227	0,566	0,395	-0,297	0,273
CALAI	-0,192	-0,002	0,014	1	0,438	0,402	0,102	0,657	0,540	0,357	0,131		0,382	0,454	0,369	-0,357	-0,221	-0,737	0,450
CHERB	0,415	0,010	0,474	0,438	1	0,909	-0,034	0,293	-0,121	0,153	0,227	0,680	0,160	0,278	0,042	0,323	0,275	-0,296	0,293
CONCA	0,536	-0,514	0,563	0,402	0,909	1		0,508	0,127	0,775	-0,214	0,850	0,319	0,416	0,285	0,677	0,078	-0,993	0,509
DIEPP	0,168		0,069	0,102	-0,034		1	-0,054	0,905	0,150	0,417		0,846	0,951	0,780	-0,068	0,171		0,849
DUNKE	-0,220	0,486	-0,174	0,657	0,293	0,508	-0,054	1	0,916	0,354	-0,101	0,888	0,425	0,663	0,792	-0,444	-0,132	-0,448	0,754
EYRAC	0,698	-0,400	0,067	0,540	-0,121	0,127	0,905	0,916	1	0,567	0,449	0,663	0,980	0,951	0,956	-0,171	0,394		0,898
HAVRE	0,681	0,551	-0,053	0,357	0,153	0,775	0,150	0,354	0,567	1	0,187	0,941	0,731	0,355	0,761	0,537	0,097		0,424
LCONQ	0,195	-0,370	0,722	0,131	0,227	-0,214	0,417	-0,101	0,449	0,187	1	0,356	0,367	0,466	0,513	0,462	0,133	0,291	0,428
LCROU	0,783	-0,091	0,475		0,680	0,850		0,888	0,663	0,941	0,356	1	0,739	0,788	0,746	0,462	0,360		0,854
LROCH	0,268	-0,671	0,288	0,382	0,160	0,319	0,846	0,425	0,980	0,731	0,367	0,739	1	0,976	0,914	0,063	0,246	-0,618	0,943
OLONE	0,548	-0,695	0,293	0,454	0,278	0,416	0,951	0,663	0,951	0,355	0,466	0,788	0,976	1	0,933	0,105	0,122	-0,865	0,961
PBLOB	0,655	0,204	0,227	0,369	0,042	0,285	0,780	0,792	0,956	0,761	0,513	0,746	0,914	0,933	1	-0,124	0,192		0,903
RSCOF	0,578	0,764	0,566	-0,357	0,323	0,677	-0,068	-0,444	-0,171	0,537	0,462	0,462	0,063	0,105	-0,124	1	-0,137	0,268	0,128
SJLUZ	0,239	0,176	0,395	-0,221	0,275	0,078	0,171	-0,132	0,394	0,097	0,133	0,360	0,246	0,122	0,192	-0,137	1		0,216
SMALO	-0,415		-0,297	-0,737	-0,296	-0,993		-0,448			0,291		-0,618	-0,865		0,268		1	
SNAZA	0,177	0,216	0,273	0,450	0,293	0,509	0,849	0,754	0,898	0,424	0,428	0,854	0,943	0,961	0,903	0,128	0,216		1

Tableau 13 : Coefficients de corrélation des événements extrêmes – Atlantique et Manche

	BOUCA	BOULO	BREST	CALAI	CHERB	CONCA	DIEPP	DUNKE	EYRAC	HAVRE	LCONQ	LCROU	LROCH	OLONE	PBLOC	RSCOF	SJLUZ	SMALO	SNAZA
BOUCA	51	13	38	13	18	10	14	14	8	18	25	6	18	15	10	24	30	5	16
BOULO	13	30	20	9	16	5	0	14	4	17	12	3	7	6	5	13	11	2	13
BREST	38	20	195	18	41	14	30	20	9	30	47	14	32	25	19	39	42	6	20
CALAI	13	9	18	43	10	6	8	27	3	14	9	2	8	7	5	8	8	5	4
CHERB	18	16	41	10	46	9	17	10	6	24	28	8	16	13	10	27	20	3	18
CONCA	10	5	14	6	9	15	2	5	6	7	11	9	12	12	9	9	5	3	7
DIEPP	14	0	30	8	17	2	47	10	4	15	15	2	9	3	6	15	21	0	15
DUNKE	14	14	20	27	10	5	10	60	4	14	11	5	10	7	8	8	11	6	7
EYRAC	8	4	9	3	6	6	4	4	15	7	8	6	7	7	7	7	7	0	6
HAVRE	18	17	30	14	24	7	15	14	7	54	19	5	10	13	7	20	12	2	11
LCONQ	25	12	47	9	28	11	15	11	8	19	50	11	22	18	9	35	28	3	16
LCROU	6	3	14	2	8	9	2	5	6	5	11	14	10	10	8	8	7	2	4
LROCH	18	7	32	8	16	12	9	10	7	10	22	10	38	15	13	18	20	4	12
OLONE	15	6	25	7	13	12	3	7	7	13	18	10	15	26	9	16	12	4	6
PBLOB	10	5	19	5	10	9	6	8	7	7	9	8	13	9	30	8	12	0	7
RSCOF	24	13	39	8	27	9	15	8	7	20	35	8	18	16	8	44	25	3	15
SJLUZ	30	11	42	8	20	5	21	11	7	12	28	7	20	12	12	25	50	0	16
SMALO	5	2	6	5	3	3	0	6	0	2	3	2	4	4	0	3	0	20	0
SNAZA	16	13	20	4	18	7	15	7	6	11	16	4	12	6	7	15	16	0	28

Tableau 14 : Nombre d'évènements extrêmes simultanés – Atlantique et Manche

## Résultats remarquables :

La qualité du coefficient de corrélation n'est pas liée à la proximité géographique des stations. Les résultats du calcul de la corrélation entre les événements extrêmes mesurés le long de la côte atlantique sont donnés à titre indicatif afin d'éventuellement être exploités lors de la mise en place de la méthode. Cependant certains résultats sont à mettre en avant :

- le plus haut coefficient de corrélation correspond au couple St Malo – Concarneau (0,993 pour 3 dates communes). Ces stations n'étant pas proches géographiquement et la quantité de dates communes étant petite, ce résultat n'apparaît pas comme étant très significatif ;
- la corrélation entre les stations de Boucau et St Jean de Luz est mauvaise (0,239 pour 30 dates communes), alors que ces stations sont très proches. La meilleure corrélation de Boucau est avec la station du Crouesty (0,783 pour 6 dates), et pour St Jean de Luz il s'agit de Brest (0,395 pour 42 dates) ;
- la station la plus corrélée avec Brest est la station du Conquet (0,722 pour 47 dates) ; cette corrélation étant aussi la meilleure pour la station du Conquet. Ces deux stations sont très proches géographiquement ;
- les stations de la Rochelle et des Sables-d'Olonne sont bien corrélées (0,976 pour 15 dates). Ces stations sont particulièrement intéressantes car elles sont proches géographiquement et toutes les deux concernées par la tempête Xynthia ;
- la plupart des coefficients de corrélation élevés pour des couples de stations éloignées correspondent à des données où peu de dates communes ont été observées (Dieppe – Les Sables-d'Olonne : 0,951 pour 3 dates ; Dunkerque - Eyrac : 0,916 pour 4 dates ; Le Havre – Le Crouesty : 0,941 pour 5 dates).

## 2. Corrélation des phénomènes extrêmes avec les autres événements

Nous allons maintenant calculer la corrélation entre les valeurs extrêmes relevées sur une station, et les données mesurées aux dates correspondantes sur les autres stations.

	AJACC	MARSE	MONAC	NICE1	PVEND	SETE1	TOULO
AJACC	1	0,471	0,589	0,554	0,766	0,235	0,538
MARSE	0,430	1	-0,071	-0,056	0,387	-0,270	-0,254
MONAC	0,739	0,403	1	0,989	0,717	0,606	0,728
NICE1	0,776	0,161	0,989	1	0,622	0,020	0,717
PVEND	0,967	0,989	0,092	0,973	1	-0,832	0,897
SETE1	-0,328	-0,069	-0,118	-0,206	-0,082	1	-0,079
TOULO	0,447	-0,024	0,551	0,861	0,700	0,106	1

Tableau 15 : Coefficients de corrélation des événements extrêmes avec tous les phénomènes des autres stations – Méditerranée

	AJACC	MARSE	MONAC	NICE1	PVEND	SETE1	TOULO
AJACC	14	14	14	12	9	13	14
MARSE	13	18	18	18	6	16	17
MONAC	10	11	23	11	6	11	19
NICE1	13	16	16	17	8	16	17
PVEND	4	3	4	3	5	3	4
SETE1	12	14	14	12	6	19	19
TOULO	13	11	20	10	6	19	33

Tableau 16 : Nombre d'évènements extrêmes – Méditerranée

Les conclusions tirées de ces calculs sont du même type que pour les autres corrélations, à savoir qu'aucun résultat ne paraît exploitable concernant la corrélation entre les stations proches géographiquement : dans certains cas le coefficient est faible, dans d'autres de nombre de valeurs utilisées pour le calculer n'est pas assez important pour que le résultat soit considéré comme significatif.

	BOUCA	BOULO	BREST	CALAI	CHERB	CONCA	DIEPP	DUNKE	EYRAC	HAVRE	LCONQ	LCROU	LROCH	OLONE	PBLOC	RSCOF	SJLUZ	SMALO	SNAZA
<b>BOUCA</b>	1	0,313	0,233	0,018	0,344	0,523	0,469	0,060	0,457	0,341	0,235	0,588	0,288	0,477	0,506	0,323	0,187	0,016	0,444
<b>BOULO</b>	0,107	1	0,069	0,021	0,328	-0,486	0,127	0,226	-0,179	0,646	-0,010	-0,549	-0,381	0,252	-0,559	0,232	0,019	-0,628	0,157
<b>BREST</b>	0,335	0,351	1	-0,103	0,380	0,719	0,249	-0,042	0,351	0,153	0,555	0,499	0,303	0,364	0,298	0,446	0,225	-0,065	0,558
<b>CALAI</b>	0,248	0,197	-0,101	1	-0,365	-0,140	0,210	0,736	0,385	0,173	-0,010	0,425	0,327	0,075	-0,111	-0,014	-0,078	0,197	0,047
<b>CHERB</b>	0,125	0,301	0,358	-0,102	1	0,895	0,299	-0,018	0,321	0,265	0,268	0,730	0,275	0,522	0,245	0,342	-0,049	-0,166	0,435
<b>CONCA</b>	-0,010	0,193	0,554	0,108	0,653	1		0,215	0,339	0,611	0,210	0,803	0,383	0,367	0,366	0,514	-0,284	-0,233	0,552
<b>DIEPP</b>	0,231	-0,330	0,162	0,265	0,127	0,951	1	0,204	0,747	0,249	0,289	0,991	0,631	0,928	0,540	0,145	0,230	0,273	0,633
<b>DUNKE</b>	0,030	0,276	-0,054	0,526	-0,263	-0,532	0,191	1	0,297	0,292	-0,070	-0,065	0,079	-0,123	-0,485	-0,018	-0,120	0,140	-0,066
<b>EYRAC</b>	-0,436	-0,322	-0,355	0,461	-0,370	-0,408	0,629	-0,349	1	-0,415	-0,290	-0,453	-0,532	-0,404	-0,389	-0,374	0,345	0,183	-0,410
<b>HAVRE</b>	0,280	0,350	0,110	0,211	0,154	0,228	0,118	0,243	0,609	1	-0,080	0,771	0,263	0,178	0,184	0,267	0,216	-0,218	0,388
<b>LCONQ</b>	0,342	-0,213	0,616	-0,094	0,160	0,309	0,305	-0,150	0,587	0,271	1	0,440	0,443	0,468	0,448	0,059	0,278	0,092	0,430
<b>LCROU</b>	0,308	-0,389	0,475	0,088	0,676	0,851		0,328	0,747	0,888	0,367	1	0,755	0,823	0,771	0,379	0,012	-0,255	0,883
<b>LROCH</b>	0,272	-0,046	0,241	0,282	-0,063	0,267	0,499	0,495	0,548	0,581	0,335	0,711	1	0,714	0,368	0,118	0,210	-0,042	0,736
<b>OLONE</b>	0,493	-0,350	0,315	0,141	0,358	0,447	0,631	0,486	0,925	0,331	0,467	0,746	0,840	1	0,910	0,330	0,256	-0,092	0,961
<b>PBLOB</b>	0,236	-0,053	0,298	0,154	0,250	0,331	0,517	0,330	0,379	0,288	0,332	0,581	0,768	0,654	1	0,230	0,257	0,036	0,769
<b>RSCOF</b>	0,375	0,642	0,549	-0,202	0,393	0,761	0,298	-0,126	0,198	0,530	0,113	0,624	0,182	0,237	0,469	1	0,134	-0,050	0,421
<b>SJLUZ</b>	0,126	0,217	0,412	0,239	0,150	0,535	-0,182	0,213	0,534	0,159	0,413	0,453	0,295	0,205	0,396	0,139	1	-0,038	0,466
<b>SMALO</b>	-0,104	0,203	0,122	0,072	-0,025	-0,048		-0,036	0,088	0,118	0,202	-0,087	0,159	0,206	0,039	0,130	0,299	1	0,978
<b>SNAZA</b>	0,139	-0,093	0,250	0,206	0,236	0,509	0,379	0,411	0,876	0,347	0,261	0,854	0,931	0,955	0,907	0,154	0,128	-0,177	1

Tableau 17 : Coefficients de corrélation des événements extrêmes avec tous les phénomènes des autres stations – Atlantique et Manche

	BOUCA	BOULO	BREST	CALAI	CHERB	CONCA	DIEPP	DUNKE	EYRAC	HAVRE	LCONQ	LCROU	LROCH	OLONE	PBLOC	RSCOF	SJLUZ	SMALO	SNAZA
<b>BOUCA</b>	51	32	51	43	47	17	30	38	19	48	46	12	29	28	18	46	42	22	28
<b>BOULO</b>	28	30	30	24	29	6	21	25	10	30	28	4	12	9	8	29	27	7	23
<b>BREST</b>	69	40	195	66	80	22	60	77	25	80	74	20	50	50	37	73	70	41	35
<b>CALAI</b>	34	18	43	43	31	16	23	35	14	35	34	10	26	24	22	32	36	18	18
<b>CHERB</b>	38	27	46	31	46	13	26	40	13	43	41	11	24	24	15	42	34	20	26
<b>CONCA</b>	15	11	15	14	15	15	2	15	11	15	14	10	15	15	12	15	8	15	8
<b>DIEPP</b>	34	27	47	30	33	3	47	42	7	30	32	3	17	4	18	32	39	3	32
<b>DUNKE</b>	41	32	60	48	41	17	42	60	17	48	46	12	34	20	28	44	48	18	33
<b>EYRAC</b>	15	10	15	9	15	11	6	13	15	15	15	12	12	13	12	15	11	12	10
<b>HAVRE</b>	44	29	54	40	51	15	27	40	12	54	51	9	29	33	19	50	39	23	23
<b>LCONQ</b>	42	27	50	34	50	15	24	37	15	49	50	16	31	31	19	50	38	27	25
<b>LCROU</b>	11	8	14	10	14	10	2	12	10	14	14	14	13	14	10	14	10	14	5
<b>LROCH</b>	28	21	38	31	32	15	18	36	12	32	31	12	38	21	20	32	26	21	21
<b>OLONE</b>	21	12	26	16	26	13	4	18	12	26	25	11	19	26	11	26	16	21	8
<b>PBLOB</b>	21	16	30	22	18	16	18	28	13	20	17	11	21	16	30	17	19	15	13
<b>RSCOF</b>	36	26	44	28	44	11	25	34	13	41	42	11	23	23	16	44	35	20	26
<b>SJLUZ</b>	42	26	50	39	40	8	38	37	13	38	39	9	26	18	21	39	50	13	29
<b>SMALO</b>	16	7	20	14	20	10	2	17	11	20	19	12	16	20	10	17	10	20	3
<b>SNAZA</b>	28	24	28	21	26	7	22	26	7	25	25	4	14	7	8	27	25	7	28

Tableau 18 : Nombre d'évènements extrêmes – Atlantique et Manche

### III. Données HYDRO

#### A. Sélection des données HYDRO

Nous nous intéressons uniquement aux stations situées sur des cours d'eau d'ordre 1, à proximité des côtes. La CCR nous a fourni la liste des cours d'eau d'ordre 1 à prendre en compte, ainsi que la liste des stations situées sur des cours d'eau d'ordre 1.

Nous avons sélectionné sous Arcgis les stations des cours d'eau intégrés aux outils de la CCR ; pour chaque cours d'eau, nous retenons la station la plus proche de la côte. La distance à la côte varie selon les cours d'eau et peut atteindre 20 à 30 km. Dans ce cas, il est fort probable qu'une submersion marine ne soit pas visible sur la hauteur du cours d'eau au niveau de la station. Nous avons tout de même ajouté ces stations à la sélection ; une étude plus approfondie des données nous permettra de les écarter. Une soixantaine de stations ont ainsi été sélectionnées.

#### B. Collecte et mise en forme des données HYDRO

Une fois la liste des stations obtenues, nous avons procédé à l'extraction des données. Les données de la Banque HYDRO peuvent être récupérées de deux manières : soit directement sur le site Internet <http://www.hydro.eaufrance.fr/>, soit par l'intermédiaire du logiciel HYDRO2 installé dans les locaux de la CCR. Nous avons opté pour la seconde solution moins coûteuse en temps.

Nous nous intéressons aux données de hauteur ; elles sont données par l'indicateur de hauteurs maximales mensuelles, ce qui est cohérent avec le choix que nous avons fait pour décrire les données SHOM. L'objectif est de déterminer, parmi les stations HYDRO, lesquelles sont susceptibles de mesurer les conséquences d'une submersion marine à proximité. Nous confrontons donc les données relevées aux données SHOM.

Pour certaines stations de la sélection, aucune donnée n'est disponible ; pour d'autres, seules des données de débit sont disponibles. Dans ce cas, nous avons remplacé, lorsqu'elles sont disponibles, les données de la station sélectionnée par celles de la station la plus proche, située sur le même cours d'eau.

Nous disposons ainsi des hauteurs maximales mensuelles de 44 stations de cours d'eau d'ordre 1 situées à proximité de côtes. Parmi elles, 40 sont des stations appartenant à la sélection initiale, et 4 stations de remplacement.

Ces données ont été mises sous forme matricielle dans un fichier Excel : chaque ligne correspond à une date (mois et année) et chaque colonne à une station.

### C. Disponibilité des données HYDRO

La première étape de l'analyse de ces données a été d'étudier leur disponibilité : étendue de l'historique et pourcentage de données manquantes. Les résultats sont donnés dans le tableau page suivante.

Code HY-DRO	Début mesures	Fin mesures	Nb données (non vides)	Nb données théorique	% données disponibles sur historique	% données disponibles 1970-2010	Remarques
E4035710	janv-68	nov-10	504	515	98%	100%	
E5400310	janv-82	nov-10	347	347	100%	71%	
G0402020	janv-99	déc-10	134	144	93%	27%	
G2011010	janv-84	déc-10	279	324	86%	57%	
I4032010	janv-71	nov-10	471	479	98%	96%	
I5231010	janv-93	sept-10	201	213	94%	41%	
I8032020	janv-93	déc-05	140	156	90%	29%	
I9221020	janv-90	oct-10	239	250	96%	49%	
J1313010	janv-79	nov-10	372	383	97%	76%	
J1513020	janv-94	nov-10	167	203	82%	34%	
J1721720	janv-84	nov-10	269	323	83%	55%	
J1813010	janv-73	nov-10	454	455	100%	92%	
J2023010	janv-81	nov-10	349	359	97%	71%	
J2034010	janv-84	nov-10	317	323	98%	65%	
J2233020	janv-93	nov-10	211	215	98%	43%	
J2723010	janv-69	oct-10	455	502	91%	93%	
J3204030	févr-10	nov-10	10	11	91%	2%	
J3323020	janv-77	nov-10	397	407	98%	81%	
J5712130	janv-88	juil-10	177	271	65%	36%	Présence de données négatives
N2013010	janv-98	nov-10	146	155	94%	30%	
Y0284060	janv-85	déc-10	312	312	100%	64%	
Y0474010	janv-05	oct-10	68	70	97%	14%	
Y0664060	janv-05	nov-10	69	71	97%	14%	
Y0824010	janv-70	nov-10	422	491	86%	86%	
Y1612020	janv-87	nov-10	226	287	79%	46%	Présence de données négatives
Y2332015	janv-07	nov-10	47	47	100%	10%	Présence de données négatives
Y2584020	janv-06	nov-10	49	59	83%	10%	Présence de données négatives
Y3204040	janv-08	nov-10	34	35	97%	7%	
Y4624010	janv-70	nov-10	480	491	98%	98%	
Y5312010	janv-70	oct-10	468	490	96%	95%	
Y5325010	janv-70	déc-10	449	492	91%	91%	Présence de données négatives
Y5444010	janv-74	nov-10	432	443	98%	88%	
Y5534030	janv-70	déc-10	480	492	98%	98%	

						Présence de données négatives
Y5615030	janv-80	déc-10	369	372	99%	75%
Y5625020	janv-87	mai-10	253	280	90%	52%
Y6442010	janv-86	déc-10	219	300	73%	45%
Y7212010	janv-80	déc-10	332	372	89%	68%
Y7915010	janv-96	déc-10	175	180	97%	36%
Y9315010	janv-60	déc-10	567	612	93%	100%
Y9605230	janv-88	déc-10	260	276	94%	53%
I0211010 (rplcmt I0211020)	janv-85	déc-10	211	312	68%	43%
I2051040 (rplcmt I2211030)	janv-93	oct-10	204	214	95%	42%
J4623020 (rplcmt J4623010)	janv-92	nov-10	224	227	99%	46%
Y4424040 (rplcmt Y4424050)	janv-97	déc-10	158	168	94%	32%

Tableau 19 : Taille de l'historique et pourcentage de données manquantes pour les 44 stations HYDRO disponibles

Pour un tiers des stations, le début d'historique est situé entre 1960 et 1980

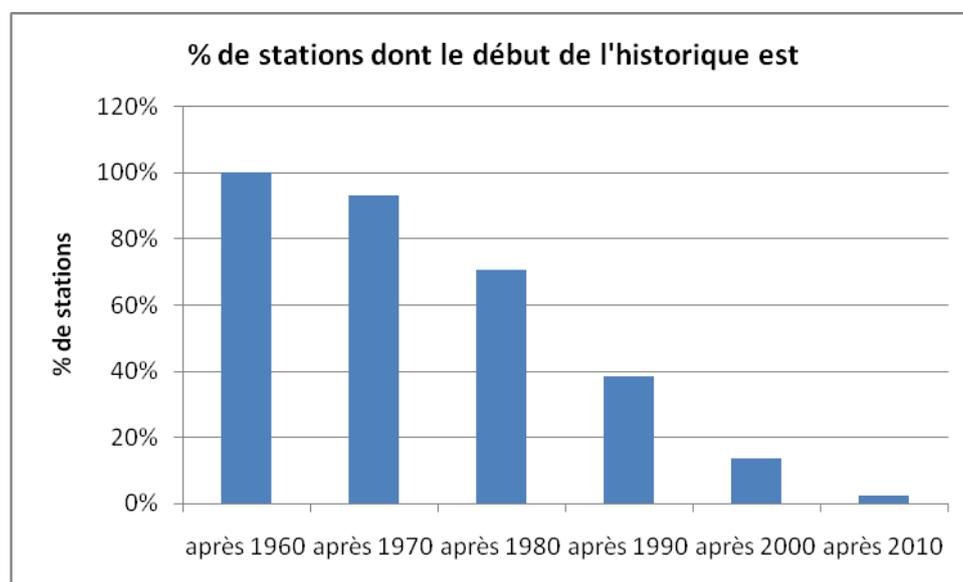


Figure 16 : Qualité des historiques disponibles : répartition des dates de début de relevé

On constate qu'en termes de taille d'historique, les données HYDRO ont une plage moins étendue que celle de certaines données SHOM. Toutefois, leurs historiques sont bien mieux complétés : la grande majorité des stations possèdent un historique complet à plus de 75% (43 stations), ce qui est loin d'être le cas des données SHOM (seulement 6 stations remplissent ce critère).

Nombre de stations dont le pourcentage de données manquantes est compris entre	
0-25%	0
25-50%	0
50-75%	3
75-100%	41

Tableau 20 : Qualité des historiques disponibles : répartition des % de données

Même lorsqu'on s'intéresse à la disponibilité des données à partir de 1970, plus de la moitié des stations ont un historique complet à plus de 50%.

Nombre de stations dont le pourcentage de données manquantes est compris entre	
0-25%	6
25-50%	15
50-75%	9
75-100%	14

Tableau 21 : Qualité des historiques disponibles : répartition des % de données

Toutefois, il faut noter que les données de certaines stations HYDRO contiennent des données négatives ; ce point doit être éclairci.

#### *D. Etude de la corrélation aux données SHOM*

##### **1. Emplacement des stations SHOM et HYDRO**

Le positionnement géographique des stations SHOM et HYDRO est représenté sur la figure ci-dessous. Nous faisons deux remarques préliminaires :

- peu de stations HYDRO sont situées sur des rivières dont l'estuaire est proche des stations SHOM ;
- les stations HYDRO sont généralement positionnées à plusieurs km à l'intérieur des terres.

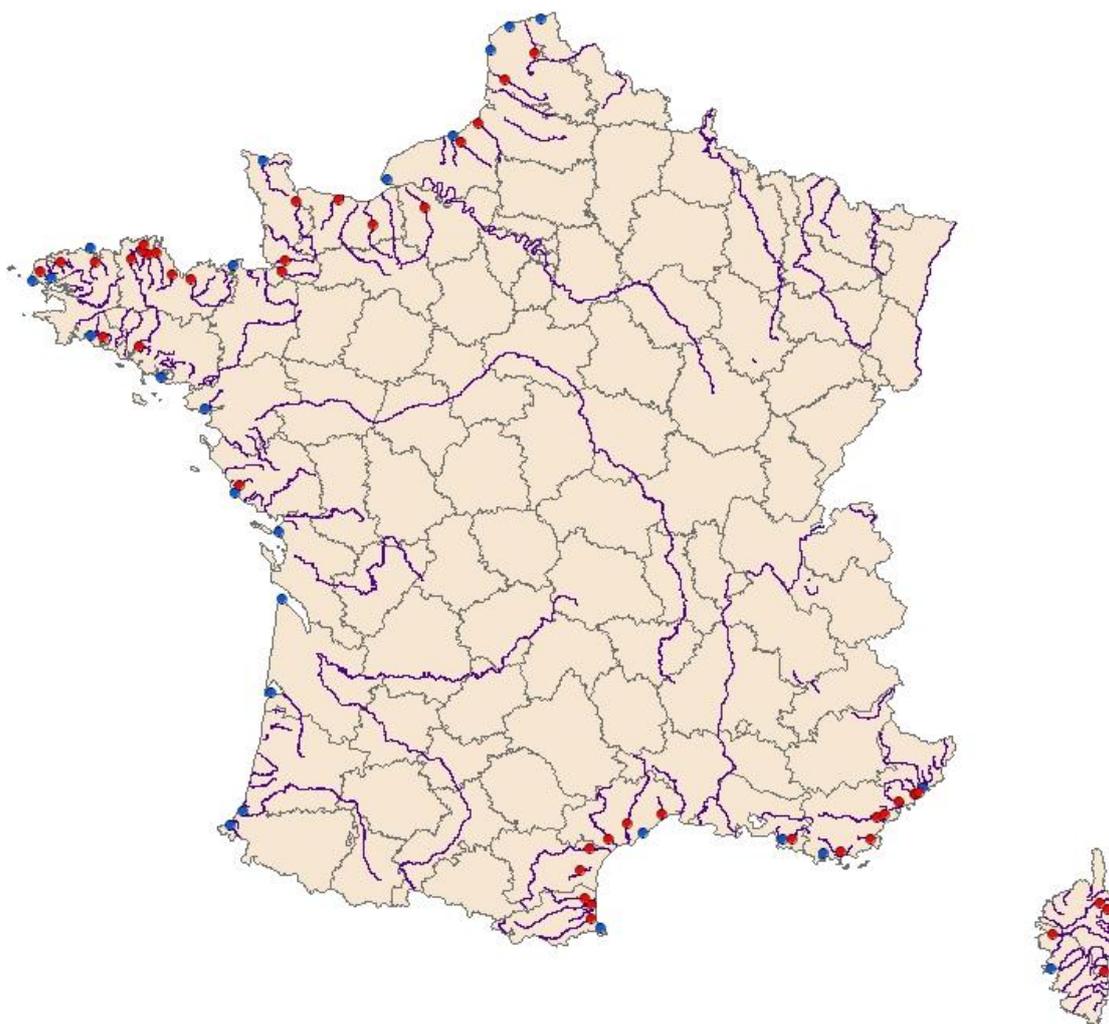


Figure 17 : Emplacement des stations SHOM et des stations HYDRO

## 2. Corrélation sur tout l'historique

Pour chaque station HYDRO, nous avons calculé le coefficient de corrélation de l'historique avec celui de chaque station SHOM, sans a priori sur la proximité station HYDRO – station SHOM.

HYDRO \ SHOM	AJACC	MARSE	MONAC	NICE1	PVEND	SETE1	TOULO
Y0284060	0,159	0,158	0,131	0,150	-0,017	0,058	0,218
Y0474010	0,055	0,133	0,029	-0,002	-0,073	0,019	0,036
Y0664060	-0,020	0,248	0,060	0,077	0,098	0,065	0,067
Y0824010	0,248	0,338	0,104	0,179	-0,030	0,088	0,213
Y1612020	0,035	0,340	0,107	0,139	0,001	0,027	0,177
Y2584020	0,170	0,273	0,243	0,231	0,196	0,100	0,281
Y2332015	0,366	0,394	0,236	0,196	0,127	0,031	0,234
Y3204040	0,452	0,249	0,279	0,236	0,080	0,069	0,260
Y4624010	0,382	0,358	0,346	0,289	0,198	-0,010	0,265

<b>Y5444010</b>	0,430	0,435	0,340	0,304	0,351	0,013	0,328
<b>Y5312010</b>	0,300	0,343	0,076	0,268	0,106	-0,004	0,231
<b>Y5325010</b>	0,118	0,204	0,127	0,162	0,105	0,091	0,141
<b>Y5534030</b>	0,274	0,303	0,212	0,265	0,322	0,000	0,273
<b>Y5615030</b>	0,298	0,296	0,231	0,236	0,426	0,014	0,262
<b>Y5625020</b>	0,194	0,241	0,247	0,258	0,235	-0,026	0,210
<b>Y6442010</b>	0,161	0,153	0,138	0,177	0,090	-0,087	0,254
<b>Y7915010</b>	0,260	0,342	0,262	0,270	0,245	0,236	0,272
<b>Y9605230</b>	0,328	0,195	0,282	0,268	0,123	0,035	0,228
<b>Y9315010</b>	0,142	0,181	-0,088	0,039	-0,121	0,018	0,037
<b>Y7212010</b>	0,330	0,291	0,096	0,285	0,121	0,059	0,264
<b>Y4424040</b>	0,316	0,292	0,291	0,286	0,228	0,181	0,223

Tableau 22 : Corrélation HYDRO - SHOM : résultats en Méditerranée

Les résultats sont mauvais : le coefficient de corrélation maximum est inférieur à 0.5. Pourtant, plus de la moitié des couples (station HYDRO, station SHOM) possède un historique commun de plus de 10 ans.

Il n'est donc pas possible d'utiliser ces résultats pour définir des couples pertinents.

Sur la côte Atlantique, les résultats sont similaires (voir tableau page suivante) : on note un coefficient élevé (-0.809), mais il s'agit de stations éloignées l'une de l'autre, et dont l'historique commun n'est constitué que de 5 dates. Les autres coefficients ne dépassent pas les 0.5.

HYDRO \ SHOM	BOUCA	BOULO	BREST	CALAI	CHERB	CONCA	DIEPP	DUNKE	EYRAC	HAVRE	LCONQ	LCROU	LROCH	OLONE	PBLOC	RSCOF	SJLUZ	SMALO	SNAZA
E4035710	0,092	0,087	0,134	0,200	0,082	0,239	0,050	0,158	0,097	0,219	0,092	-0,108	0,081	0,062	0,189	0,085	0,032	0,100	0,079
E5400310	0,179	0,106	0,104	0,098	0,104	0,154	0,122	0,147	-0,031	0,201	0,024	-0,046	0,090	0,076	0,208	0,046	-0,004	0,118	0,011
G0402020	0,254	0,001	0,214	0,116	0,160	0,258	-0,102	0,096	0,034	0,208	0,152	-0,151	0,119	0,115	0,232	0,169	0,005	0,135	0,081
G2011010	0,246	0,134	0,199	0,180	0,208	0,290	0,038	0,275	0,097	0,287	0,154	-0,059	0,139	0,170	0,271	0,134	0,127	0,116	0,012
I4032010	0,143	0,188	0,164	0,171	0,135	0,254	0,118	0,157	0,052	0,265	0,158	-0,008	0,151	0,092	0,233	0,100	0,051	0,092	0,046
I5231010	0,150	0,153	0,190	0,136	0,205	0,292	-0,194	0,382	-0,016	0,337	0,153	0,030	0,180	0,161	0,275	0,120	-0,053	0,032	0,133
I8032020	0,162	0,096	0,149	0,033	0,150	0,283	0,038	0,267	-0,053	0,292	0,118	-0,181	0,132	0,036	0,317	0,046	-0,088	0,038	0,145
I9221020	0,162	0,156	0,184	0,154	0,194	0,231	-0,296	0,342	-0,031	0,269	0,168	-0,060	0,140	0,121	0,179	0,106	-0,014	0,063	-0,059
J1313010	0,135	0,195	0,103	0,163	0,136	0,199	0,078	0,098	0,040	0,192	0,090	0,060	0,095	0,098	0,187	0,078	0,006	0,022	0,053
J1513020	0,184	0,119	0,226	0,234	0,221	0,285	-0,187	0,304	0,041	0,349	0,203	0,051	0,215	0,216	0,284	0,168	-0,036	0,081	0,176
J1813010	0,227	0,262	0,203	0,230	0,152	0,215	0,112	0,232	0,071	0,299	0,175	0,069	0,212	0,157	0,278	0,137	0,098	-0,013	0,146
J1721720	0,139	0,132	0,138	0,215	0,145	0,262	-0,301	0,280	-0,009	0,248	0,141	0,051	0,154	0,130	0,309	0,051	-0,062	0,079	0,099
J2023010'	0,170	0,210	0,141	0,201	0,161	0,233	0,053	0,207	0,048	0,241	0,126	0,023	0,164	0,134	0,320	0,085	0,002	0,047	0,075
J2034010	0,138	0,188	0,133	0,136	0,152	0,250	0,039	0,254	0,074	0,233	0,108	0,047	0,194	0,145	0,358	0,085	-0,002	0,025	-0,014
J2233020	0,163	0,120	0,136	0,211	0,149	0,232	-0,171	0,250	-0,018	0,237	0,132	-0,008	0,159	0,118	0,302	0,061	-0,033	0,040	0,115
J2723010	0,153	0,208	0,201	0,290	0,123	0,158	0,198	0,226	0,032	0,315	0,170	0,048	0,160	0,114	0,223	0,122	0,096	0,143	0,089
J3204030	0,441	-0,132	-0,055	-0,107	0,053	-0,066	-0,210	-0,020	0,088	-0,073	-0,097	-0,113	-0,079	-0,012	-0,020	0,020	0,237	-0,809	-0,142
J3323020	0,204	0,223	0,188	0,190	0,175	0,229	0,136	0,224	0,074	0,297	0,177	0,100	0,151	0,159	0,258	0,133	0,114	0,105	0,076
J5712130	0,072	0,041	0,051	0,128	0,080	0,160	-0,380	0,138	-0,025	0,187	0,005	-0,074	0,069	0,005	0,209	-0,039	-0,181	0,020	-0,003
N2013010	0,222	0,223	0,231	0,339	0,263	0,331	0,165	0,286	0,029	0,328	0,185	0,093	0,199	0,166	0,245	0,176	0,066	0,319	0,156
I0211010	0,057	-0,012	0,023	-0,032	-0,006	0,174	-0,114	0,134	-0,055	0,089	-0,055	-0,110	0,170	0,031	0,279	-0,014	-0,055	0,105	-0,120
I2051040	0,118	0,069	0,068	0,018	0,075	0,184	-0,359	0,229	-0,039	0,193	0,082	-0,020	0,113	0,063	0,236	-0,003	-0,022	0,007	0,116
J4623020	0,187	0,113	0,215	0,149	0,216	0,287	-0,050	0,295	0,007	0,269	0,188	0,091	0,189	0,158	0,315	0,121	0,032	0,077	0,098

Tableau 23 : Corrélation HYDRO - SHOM : résultats pour la côte Atlantique

### E. Etude de la corrélation des évènements extrêmes

L'étude des corrélations des historiques des données HYDRO à ceux des données SHOM ne suffit pas à déterminer si certaines stations mesurent l'impact d'une immersion côtière sur la hauteur des cours d'eau associés.

Nous complétons donc cette étude par l'analyse conjointe des évènements extrêmes HYDRO – SHOM. Pour ce faire, nous nous intéressons aux dates pour lesquelles la hauteur relevée est considérée comme susceptible de provoquer une inondation ou une submersion : nous faisons l'hypothèse que cela se produit lorsque la hauteur mesurée est supérieure au quantile à 90% des hauteurs contenues dans l'historique. Une fois ces dates sélectionnées, nous étudions la corrélation par couple (station HYDRO, station SHOM) sur les évènements extrêmes. Les résultats obtenus sont ensuite confrontés au nombre de dates où les deux stations connaissent simultanément une inondation : si ce nombre est faible, le résultat de la corrélation est peu fiable.

#### 1. Résultats en Méditerranée

Les résultats obtenus en Méditerranée sont donnés dans le tableau suivant :

HYDRO \ SHOM	AJACC	MARSE	MONAC	NICE1	PVEND	SETE1	TOULO
Y0284060		0,68		-1		-0,60	-0,37
Y0474010		1					
Y0664060		1					
Y0824010	1	1		-1		-0,51	-0,33
Y1612020		1				-1	
Y2584020	-0,84	-1	-1	-1			-1
Y2332015	-1	1					
Y3204040							
Y4624010	0,23	-0,15	0,54	-0,50			0,005
Y5444010	-1	-0,43		-1		-0,72	-0,08
Y5312010		-0,94					-0,95
Y5325010		-0,22					
Y5534030		0,64				1	1
Y5615030	1	1				1	1
Y5625020		0,27					
Y6442010		1,00				-1,00	
Y7915010	-0,94	0,13	-1	-0,67		1	-1
Y9605230						0,78	0,87
Y9315010		1		1		1	-1
Y7212010	-0,82	0,51		-0,25		0,94	
Y4424040	0,09	-0,52		-0,85		1	1

Tableau 24 : Corrélation HYDRO – SHOM sur les évènements extrêmes : résultats en Méditerranée

Certaines cases sont vides : cela signifie qu'il y a moins de deux dates d'évènement extrême communes aux deux stations. C'est notamment le cas pour la station Port-Vendres.

HYDRO \ SHOM	AJACC	MARSE	MONAC	NICE1	PVEND	SETE1	TOULO
Y0284060	1	4	0	2	0	4	3
Y0474010	2	2	0	0	0	0	0
Y0664060	1	2	0	0	0	0	0
Y0824010	2	2	0	2	0	4	4
Y1612020	0	2	0	1	0	2	1
Y2584020	3	2	2	2	1	1	2
Y2332015	2	2	1	1	0	0	1
Y3204040	1	1	0	0	0	0	0
Y4624010	3	7	3	4	0	1	4
Y5444010	2	6	1	2	1	3	5
Y5312010	1	3	0	0	0	1	3
Y5325010	1	3	0	0	0	1	1
Y5534030	1	3	0	1	0	2	2
Y5615030	2	2	0	1	0	2	2
Y5625020	1	3	0	1	0	1	0
Y6442010	1	2	0	1	0	2	1
Y7915010	3	5	2	4	0	2	2
Y9605230	1	1	0	1	0	3	3
Y9315010	0	2	0	2	1	2	2
Y7212010	4	3	0	3	0	3	1
Y4424040	3	6	1	3	0	2	2

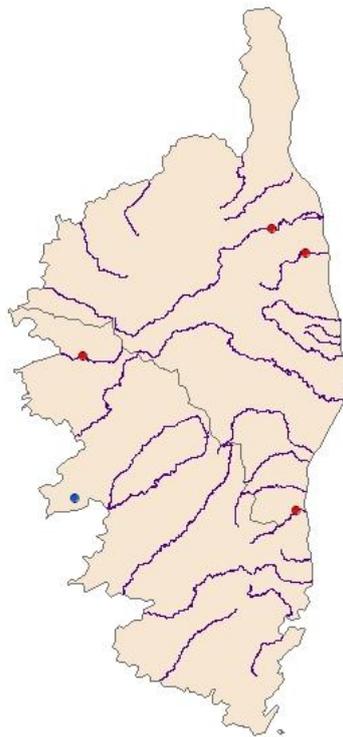
Tableau 25 : Corrélation HYDRO – SHOM sur les évènements extrêmes :  
nombre de crues simultanées en Méditerranée

On remarque un certain nombre de coefficients valant 1 ou -1; c'est le cas lorsqu'il n'y a que deux dates communes aux évènements extrêmes des stations ; les résultats ne sont donc pas pertinents.

### Résultats remarquables :

- La station Y2584020 et Ajaccio ont un coefficient de corrélation de -0,84 établi sur 3 dates communes. Or cette station HYDRO est proche des stations de SETE et PORT VENDRES, mais pas d'Ajaccio ;
- La station Y0284060 a un coefficient de 0,68 avec Marseille (4 dates communes), -0,6 avec Sète (4 dates) et -0,37 avec Toulon (3 dates). La station est située près de Port-Vendres, mais on ne dispose pas d'assez de dates communes pour calculer la corrélation ;
- Y0824010 est située entre Port-Vendres et Sète, pourtant sa corrélation avec Sète est mauvaise (-0,51 pour 4 dates) ;
- Y4624010 est proche de Toulon, pourtant la corrélation entre les deux est presque nulle (0.005 pour 4 dates communes) ;

- Y5444010 est proche de Toulon et Nice, pourtant sa corrélation est de -0.08 avec Toulon (5 dates communes) et bonne avec Sète (-0.72 pour 3 dates) ;
- Y5312010 se situe entre Toulon et Nice. La corrélation avec Toulon est bonne (-0.95 pour 3 dates), mais aussi avec Marseille (- 0.94 pour 3 dates) ;
- Y7915010 est proche d'Ajaccio : leur coefficient de corrélation est bon (-0.94 pour 3 dates) ;
- Y9605230 est proche d'Ajaccio. Elle est cependant plus corrélée avec Sète et Toulon (0.78 et 0.87 pour 3 dates) ;



*Figure 18 : Emplacement des stations SHOM et HYDRO le long de la côte corse*

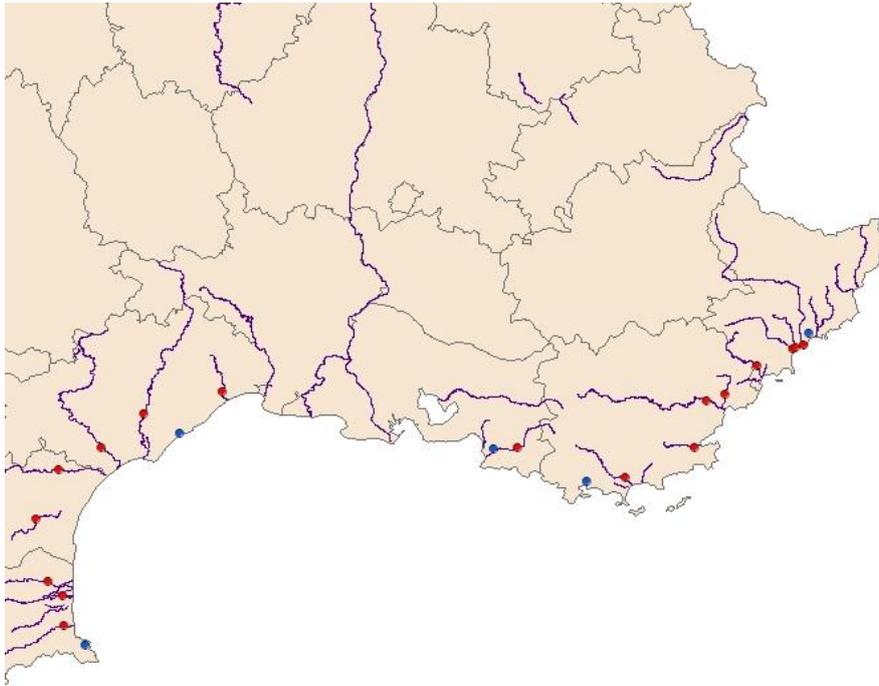


Figure 19 : Emplacement des stations SHOM et HYDRO le long de la côte méditerranéenne

En visualisant le positionnement des stations HYDRO par rapport à la côte et aux stations SHOM, seules certaines semblent exploitables. Il s'agit de :

- Y4624010 : proche de Toulon ;
- Y5534030 : proche de Monaco et Nice ;
- Y5615030 : proche de Monaco et Nice ;
- Y5625020 : proche de Monaco et Nice ;
- Y6442010 : proche de Monaco et Nice ;
- Y4424040 : proche de Marseille.

Cependant les coefficients de corrélation entre ces stations HYDRO et les stations SHOM proches sont insuffisants, il semble donc que les mesures réalisées sur les stations HYDRO le long de la côte méditerranéenne soient inexploitables.

## 2. Résultats sur la côte Atlantique et la Manche

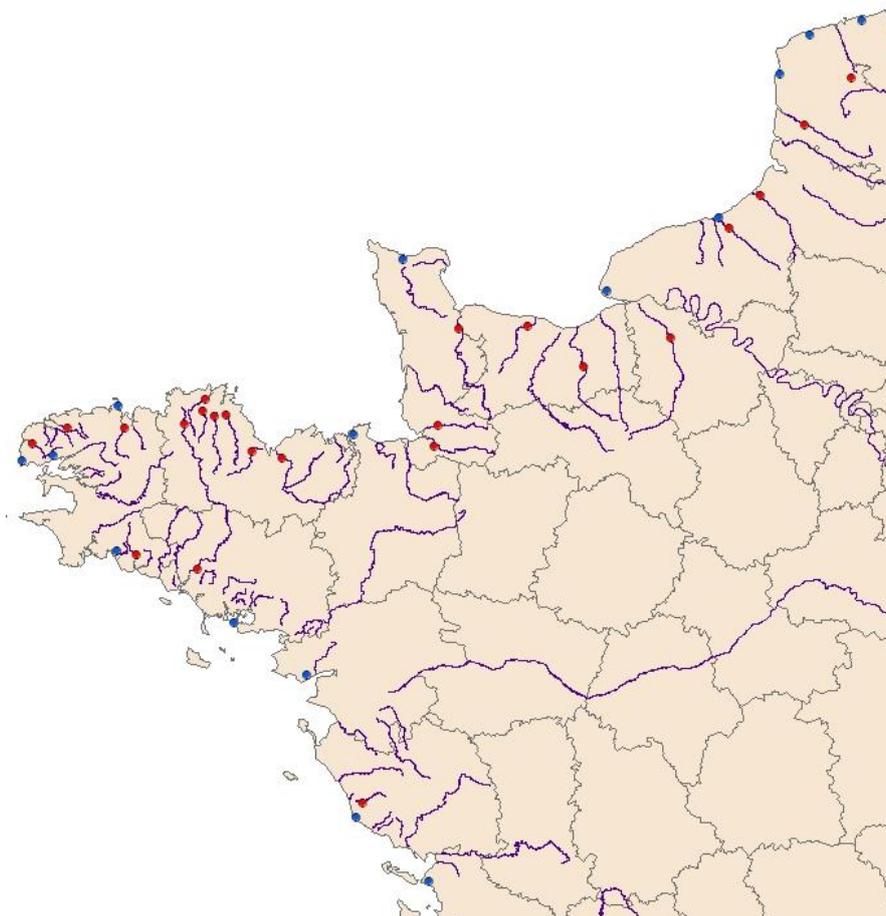
HYDRO \ SHOM	BOUCA	BOULO	BREST	CALAI	CHERB	CONCA	DIEPP	DUNKE	EYRAC	HAVRE	LCONQ	LCROU	LROCH	OLONE	PBLOC	RSCOF	SJLUZ	SMALO	SNAZA
E4035710	0,779		0,735	0,520	0,811	0,993		0,652	0,999	0,651	0,301		0,958	0,879	-0,967		1	-0,580	
E5400310	0,152	-0,844	-0,394	0,705	-0,099	0,681		-1	0,924	0,511	0,142		0,883	-0,833	0,576	0,732	1	-0,967	0,545
G0402020	0,639		-1	1	1	0,987				0,602	-1		1	1	0,167			-1	1
G2011010	0,057		-0,199	0,578	0,188	0,697		-1		-0,450	-0,555		0,943	-0,216	0,509			0,060	
I4032010	0,300	-0,503	-0,210	-0,760	-0,554	0,943		-0,482	-0,052	-0,449	-0,540		0,062	-0,407	0,959	0,560	-0,314	-0,501	-0,904
I5231010	-0,445			0,762				-0,448	-1	0,946	-1		0,768	-0,114				-1	
I8032020	-0,180			-1		-1				0,984			1	1	-1			-0,045	
I9221020	0,966			0,342	1			0,386		0,349	1		0,605	-0,676				-0,413	
J1313010	0,420	-0,273	-0,165	-0,887	-0,376	1		-0,617	-0,597	0,469	0,339				-0,655	0,812	-0,787	-0,759	0,812
J1513020	0,773			-0,808				-1	-1	0,641					-1			-0,519	
J1813010	0,200	0,029	-0,252	-0,062	-0,018	0,782		-0,132	-0,689	0,544	0,081		0,626	0,571	-0,169	0,821	-0,806	-0,438	0,260
J1721720	0,523		0,819	-0,234	-0,364	0,856		0,358		-0,184	0,519		0,989	-0,484	0,219			-0,420	
J2023010'	0,117	0,577	0,432	0,126	0,223	0,500		-0,756	-0,983	-0,307	-1		0,975	-0,672	0,179	1	-0,952	-0,306	-1
J2034010	0,667	1	0,458	-0,788				-1	-1	0,799				-0,683	-1	-1	-0,559	-1	
J2233020	0,503		0,580	-0,443		0,737		-1		0,565	0,630		1	-0,904	0,423		1	-0,618	
J2723010	0,917	1	-0,621	-0,499	1	0,979		-0,677	-1	-0,074	-0,143		-0,080	-0,596	1	1	1	-0,789	1
J3204030																			
J3323020	0,660	0,709	-0,579	0,142	-0,502	-0,063		-0,458	0,999	0,802	0,215		0,959	-0,178	0,149	1	0,181	-0,897	0,264
J5712130	0,860		1	-0,697		0,990				0,486	0,558		1	-0,992	0,893			-0,844	
N2013010	0,226		1	0,860		0,711		-0,024		-1	1		0,621	0,977				-0,428	
I0211010	-0,249	0,144			1	1		-1	1	0,033					1			-1	1
I2051040	-0,705			0,255	-1			0,790		0,383			-1	-0,277	-1			-0,770	
J4623020	0,236		1	0,705		0,849		1	-1	0,445	0,882		1	-0,995	0,171			-0,912	

Tableau 26 : Corrélation HYDRO – SHOM sur les évènements extrêmes : résultats sur la côte Atlantique et la Manche

HYDRO \ SHOM	BOU-CA	BOU-LO	BRES-T	CA-LAI	CHER-B	CON-CA	DIEP-P	DUNK-E	EY-RAC	HAVR-E	LCON-Q	LCRO-U	LROC-H	OLON-E	PBLO-C	RSCO-F	SJLU-Z	SMA-LO	SNA-ZA
E4035710	6	1	3	6	3	3	1	5	3	8	3	1	4	5	3	0	2	6	1
E5400310	7	4	5	5	4	3	0	2	3	5	5	0	3	3	4	4	2	4	3
G0402020	5	1	2	2	2	3	0	0	0	4	2	0	2	2	4	1	0	2	2
G2011010	5	1	4	4	4	5	0	2	1	6	4	1	4	4	4	1	0	3	1
I4032010	11	4	6	5	5	5	1	9	3	5	6	2	5	3	5	3	5	6	4
I5231010	4	0	1	4	1	1	1	3	2	5	2	0	3	5	1	1	1	3	0
I8032020	5	0	1	3	1	2	0	1	0	3	1	0	2	2	2	0	0	3	1
I9221020	4	0	1	5	2	1	1	6	1	3	2	0	3	3	0	0	1	6	0
J1313010	6	5	5	3	5	2	1	6	3	8	3	0	1	1	3	3	3	3	3
J1513020	4	0	1	3	2	1	1	2	2	4	1	0	1	1	2	0	1	5	0
J1813010	10	7	6	9	6	3	1	11	4	10	6	1	6	3	4	4	7	3	7
J1721720	5	1	4	5	4	3	0	4	1	6	3	1	3	3	4	1	1	4	1
J2023010'	7	4	4	7	4	3	1	6	3	7	2	1	3	3	4	2	3	4	2
J2034010	7	2	4	4	1	2	1	2	2	6	2	0	1	3	2	2	4	2	1
J2233020	6	0	3	5	2	3	1	2	1	5	3	1	2	3	4	1	2	4	1
J2723010	4	2	4	4	2	4	0	6	2	4	5	1	4	3	2	2	2	3	2
J3204030	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
J3323020	9	6	5	6	5	4	1	7	3	7	4	1	5	4	4	2	3	4	5
J5712130	5	1	2	3	2	3	0	1	1	5	3	1	2	3	3	0	0	5	1
N2013010	4	1	2	5	1	3	0	3	1	2	2	0	4	4	1	1	1	7	0
I0211010	5	4	2	1	2	2	0	2	2	4	2	0	1	1	2	1	1	2	2
I2051040	4	1	1	6	2	1	1	4	1	4	1	0	2	3	2	1	1	4	1
J4623020	5	0	2	4	2	3	0	2	2	4	3	1	2	3	3	0	0	6	1

Tableau 27 : Corrélation HYDRO – SHOM sur les événements extrêmes : nombre de crues simultanées sur la côte Atlantique et la Manche

Beaucoup de coefficients de corrélation entre les événements extrêmes des stations SHOM et HYDRO situées sur la côte atlantique et le long de la Manche apparaissent comme élevés. Cependant, le positionnement de ces stations les unes par rapport aux autres ne peut pas être ignoré. Certaines stations HYDRO sont très éloignées de la côte et apparaissent comme n'étant proches d'aucune station SHOM ; il ne semble pas que l'exploitation de ces stations soient pertinente.



*Figure 20 : Emplacement des stations SHOM et HYDRO le long des côtes bretonnes et de la Manche*



Figure 21 : Emplacement des stations SHOM et HYDRO le long de la côte Atlantique

De la même façon que pour la côte méditerranéenne, nous sélectionnons uniquement quelques stations HYDRO qui sont proches de stations SHOM. Ces stations sont :

- G2011010 : proche de Dieppe ;
- J2723010 : proche de Roscoff ;
- N2013010 : proche des Sables-d'Olonne ;
- J4623020 : proche de Concarneau.

Seules les deux dernières associations de stations sont exploitables. En effet pour la station N2013010 et la station OLONE, le coefficient de corrélation est de 0.977 pour 4 dates communes, et pour J4623020 et CONCA, il est de 0.849 pour 3 dates.

Les dates communes d'évènements extrêmes pour les stations N2013010 et OLONE sont les mois de décembre 1998, novembre 1999, mars 2001 et novembre 2006. Ces dates ne correspondent à aucun évènement recensé dans le fichier répertoriant les tempêtes ayant entraîné des Chocs Mécaniques liés aux Vagues (CMV). Ni la tempête de décembre 1999, ni la tempête Xynthia, ne sont à l'origine d'hauteurs d'eau extrêmes mesurées par la station HYDRO concernée.

*F. Corrélation entre la station SHOM de Cherbourg et la station HYDRO I6943110 (La Divette)*

La station HYDRO I6943110 qui se situe sur la Divette à Octeville est très proche de la station SHOM de Cherbourg. Nous allons donc chercher à voir si cette station HYDRO est exploitable pour mesurer l'influence de la hauteur d'eau côtière sur la hauteur d'eau de la rivière.

Ces deux stations ont 290 mesures réalisées à des dates communes, ce qui représente plus de 20 ans de mesures, soit un historique assez représentatif.

La valeur du coefficient de corrélation entre ces stations est de 0,196 pour 290 dates. Ce résultat est très faible, et fait ressortir le fait qu'aucune corrélation n'existe entre ces stations.

Cependant, afin de confirmer ce résultat, nous avons cherché à voir s'il existait des événements extrêmes communs aux deux stations. Seules deux dates répondent à ce critère :

- le mois de janvier 1995 ;
- le mois d'avril 1998.

Ces dates ne correspondent à aucun événement CMV recensé dans le fichier qui nous a été fourni.

Des inondations ont eu lieu dans sur la commune de Cherbourg en décembre 2010. Cependant, les données de la station SHOM ne sont pas encore accessibles pour ce mois-là. Nous ne pouvons pas étudier le lien entre ces inondations et la hauteur de la marée.

Cette étude confirme donc que les données HYDRO ne seront pas exploitables dans l'étude menée sur les hauteurs d'eau côtières extrêmes.

## IV. Sélection des stations exploitables

### A. Critères de sélection des stations

L'étude des phénomènes extrêmes des hauteurs de marées nécessite de choisir, sur la base de critères précis, les stations pour lesquelles la méthodologie pourra être appliquée.

La sélection des stations est effectuée suivant les quatre critères suivants classés en fonction de leur niveau de contrainte :

- Critère n°1 : pour l'étude des phénomènes extrêmes, seulement quelques stations pourront être traitées conjointement et il est exclu d'appliquer la méthodologie simultanément à toutes les stations du littoral français.
- Critère n°2 : l'objectif étant l'étude des phénomènes extrêmes, la qualité des résultats dépend directement de la taille de l'historique des stations considérées. Les stations ayant un historique commun très important sont donc retenues prioritairement. Les autres stations sont utilisées uniquement pour la reconstruction de données manquantes.
- Critère n°3 : Pour obtenir un historique commun et complet très important, il est nécessaire de reconstituer les données manquantes de chaque station. La reconstitution de données introduit des données fictives et ne permet pas la reconstruction des événements les plus extrêmes. Les stations ayant le pourcentage de données manquantes le plus faible sont donc retenues prioritairement. Les autres stations sont utilisées uniquement pour la reconstruction de données manquantes.
- Critère n°4 : Les stations restantes ayant les phénomènes extrêmes les plus marqués sont finalement retenues. Les autres stations sont utilisées uniquement pour la reconstruction de données manquantes.

### B. Exemple de sélection de deux stations

#### 1. Choix du littoral

Le critère n°1 impose de limiter le nombre de stations traitées. Sur les 26 stations disponibles sur l'ensemble du littoral français, seulement 7 concernent la Méditerranée. Ces stations étant réparties géographiquement de manière homogène, nous choisissons de traiter notre exemple avec deux stations méditerranéennes.

#### 2. Choix de l'historique

Pour respecter le critère n°2, il est nécessaire de sélectionner l'historique le plus long pouvant être utilisé pour définir la loi conjointe des stations SHOM.

Les statistiques pour les stations se situant sur la côte méditerranéenne sont :

	Mois début	Année début	Mois fin	Année fin	Nb données (théorique)	Nb données (réel)	Pourcentage données manquantes
AJACC	8	1981	6	2010	347	121	65,13%
MARSE	2	1985	4	2010	303	162	46,53%
MONAC	1	1960	6	2010	606	216	64,36%
NICE1	8	1981	6	2010	347	163	53,03%
PVEND	1	1982	6	2010	342	46	86,55%
SETE1	2	1992	6	2010	221	184	16,74%
TOULO	1	1961	6	2010	594	323	45,62%

Tableau 28 : Statistiques des stations situées sur la côte méditerranéenne

Faire débiter l'historique à traiter avant 1981 ne serait pas un choix judicieux, car seules deux stations disposent de données antérieures à cette date. Cependant, avant 1985, trop de données manquantes sont recensées. Une analyse plus approfondie des données permet de constater que de 1987 à 1992, aucune des stations n'a été relevée. Le choix est donc fait de faire débiter l'historique de notre exemple en février 1992. A cette étape, les 7 stations méditerranéennes sont conservées.

### 3. Reconstruction des données manquantes

Une fois l'historique sélectionné, une reconstruction des données manquantes est effectuée pour les stations du bassin méditerranéen. La méthode mise en place est la même que celle utilisée dans le cadre du contrat précédent passé entre la SCM et la CCR.

Les tableaux ci-dessous représentent les résultats de reconstruction pour des coefficients de corrélation supérieurs à 0,9 et 0,5.

	AJACC	MARSE	MONAC	NICE1	PVEND	SETE1	TOULO
nombre de données initiales	106	137	134	139	31	184	208
% de données manquantes initialement	52 %	38 %	40 %	37 %	86 %	17 %	6 %
nombre de données après reconstruction	106	137	210	211	159	184	208
% de données manquantes après reconstruction	52 %	38 %	5 %	5 %	28 %	17 %	6 %

Tableau 29 : Statistiques liées à la reconstruction de données manquantes pour un coefficient de corrélation supérieur à 0,9

	AJACC	MARSE	MONAC	NICE1	PVEND	SETE1	TOULO
nombre de données initiales	106	137	134	139	31	184	208
% de données manquantes initialement	52 %	38 %	40 %	37 %	86 %	17 %	6 %
nombre de données après reconstruction	221	221	221	221	215	184	221
% de données manquantes après reconstruction	0 %	0 %	0 %	0 %	3 %	17 %	0 %

Tableau 30 : Statistiques liées à la reconstruction de données manquantes pour un coefficient de corrélation supérieur à 0,5

Les stations Ajaccio et Port-Vendres possèdent 52% et 87% de données manquantes. D'après le critère n°3, ces stations ne sont pas retenues.

Plus le seuil fixé pour la reconstruction pour le coefficient de corrélation est bas, plus la quantité de données reconstruites est importante ; cependant, la reconstruction est de moins bonne qualité. Les stations nécessitant des coefficients de corrélation très faibles pour être reconstituées ne sont pas retenues : il s'agit de Marseille et Sète.

### *C. Stations exploitables*

Les critères définis ci-dessus permettent de réduire le nombre de stations pouvant être étudiées pour chacun des littoraux.

Ainsi pour la Méditerranée, seules 5 stations sont conservées :

- Ajaccio ;
- Monaco ;
- Marseille ;
- Nice ;
- Toulon.

L'historique à étudier pour ces stations est de 263 mois (environ 22 ans) divisés entre deux périodes distinctes : du mois d'août 1981 au mois de juillet 1985 et d'octobre 1991 à juin 2010.

De la même façon pour la côte atlantique et la Manche, les 4 critères de sélection des stations réduisent à 12 le nombre de stations dont les données sont considérées comme satisfaisantes.

Il s'agit de :

- Boucau ;
- Boulogne-sur-mer ;
- Brest ;
- Cherbourg ;
- Dieppe ;
- Le Havre ;
- Le Conquet ;
- La Rochelle ;
- Les Sables d'Olonne ;
- Roscoff ;
- Saint Jean-de-Luz ;
- Saint Nazaire.

L'historique conservé pour ces stations s'étend sur 474 mois (environ 39 ans) consécutifs, de janvier 1971 à juin 2010.

## V. Extension de la loi conjointe : théorie

Nous expliquons la méthode permettant d'étendre la loi conjointe aux phénomènes extrêmes.

### A. Premiers éléments

Pour simplifier, nous nous limitons au cas de trois stations de mesure. L'information disponible est un historique de mesures effectuées sur les trois stations ; il se présente sous la forme d'un tableau à trois colonnes ; on met en ligne la mesure réalisée à une date donnée :

	Station 1	Station 2	Station 3
date 1	$x11$	$x12$	$x13$
date 2	$x21$	$x22$	$x23$
...	...	...	...
date N	$xN1$	$xN2$	$xN3$

Les quantités mesurées ne sont pas nécessairement de la même nature, ni dans les mêmes unités ; ce qui est important c'est la simultanéité : à la même date  $k$ , on a mesuré  $x_{k,i}$  dans la  $i$ -ème ville.

A partir de ce tableau, on réalise un découpage des mesures pour chaque station en classes. Le nombre de classes n'a pas de raison d'être le même d'une station à l'autre ; par exemple, on peut avoir dix classes pour la station 1, douze pour la station 2 et cinq pour la station 3.

Si nous notons  $C_1$  le nombre de classes pour la station 1,  $C_2$  pour la station 2 et  $C_3$  pour la station 3, le nombre de triplets possibles, pour les résultats, est  $C_1 \times C_2 \times C_3$ . Si nous avons  $S$  stations et que le nombre de classes est égal à  $C$  pour toutes, le nombre de résultats possibles, pour la loi conjointe, est  $C^S$ . Chaque classe, comme d'habitude, correspond à une discrétisation : quelle est la précision des résultats que l'on souhaite garder ?

Dans la suite, par simplicité d'écriture, nous nous limitons à trois stations où des quantités de même nature sont observées (par exemple des hauteurs d'eau, pour fixer les idées) ; la discrétisation et le nombre de classes (noté  $C$ ) seront les mêmes pour toutes.

Les résultats possibles d'une observation conjointe seront donc tous les triplets de la forme  $(x_i, x_j, x_k)$  où  $x_i, x_j, x_k$  prennent chacun les  $C$  valeurs possibles.

Par exemple, si les résultats sont dix hauteurs possibles, de 600 à 1500 cm, en considérant un pas de 100 cm, il y a mille triplets, commençant par (600,600,600) et finissant par (1500,1500,1500).

## Notations

Il sera commode de noter  $x(i, j, k)$  le triplet correspondant à la  $i$ -ème valeur pour la première station, à la  $j$ -ème pour la seconde et à la  $k$ -ème pour la troisième. Dans l'exemple précédent,  $x(4, 2, 5) = (900, 700, 1000)$ .

Nous noterons  $n(i, j, k)$  le nombre de fois où le triplet  $x(i, j, k)$  a été rencontré dans le tableau des observations. On aura donc :

$$\sum_{i,j,k=1}^c n(i, j, k) = N$$

où  $N$  désigne le nombre total d'observations.

Enfin, nous notons  $\lambda(i, j, k)$  le "taux de risque" (probabilité) du triplet  $(i, j, k)$  : c'est la quantité inconnue qu'il s'agit d'évaluer, compte-tenu des observations. De manière grossière, on sait que :

$$\lambda(i, j, k) \sim \frac{n(i, j, k)}{N}$$

mais on veut une loi de probabilité sur  $\lambda(i, j, k)$ , et pas seulement une estimation ponctuelle. La théorie générale (voir BB [1]) dit que la loi conjointe de l'ensemble des taux de risque, notée  $f\left(\left(\lambda(i, j, k)\right)_{i,j,k}\right)$  est donnée par la formule :

$$f\left(\left(\lambda(i, j, k)\right)_{i,j,k}\right) = c 1_S \prod_{i,j,k} \lambda(i, j, k)^{n(i,j,k)} \quad (1)$$

où  $c$  est une constante de normalisation (pour que l'intégrale de  $f$  par rapport à toutes les variables soit égale à 1) et  $S$  est un simplexe que nous allons maintenant définir.

### *B. Le simplexe de travail*

Nous admettons (voir [1]) que, en un site donné, les probabilités sont décroissantes lorsque le phénomène devient de plus en plus extrême ; par exemple, s'il s'agit d'une température, rencontrer 39°C est moins probable que rencontrer 38°C. En termes de hauteurs, cela signifie qu'il est moins probable d'observer une hauteur de 1500 cm qu'une hauteur de 1400 cm. Par contre, les probabilités ne sont pas comparables s'il s'agit de stations distinctes.

Ceci se traduit par les inégalités :

$$\begin{aligned} \lambda(i, j, k) &\geq \lambda(i+1, j, k) \\ \lambda(i, j, k) &\geq \lambda(i, j+1, k) \\ \lambda(i, j, k) &\geq \lambda(i, j, k+1) \end{aligned} \quad (2)$$

Notre simplexe de travail  $S$  est donc constitué de  $C^3$  variables  $\lambda(i, j, k)$  astreintes aux conditions (1), et aux conditions :

$$\lambda(i, j, k) \geq 0$$

$$\sum_{i,j,k=1}^c \lambda(i, j, k) = 1.$$

Par exemple, si l'on dispose de deux stations et de deux classes seulement, la fonction  $f$  ci-dessus dépend des quatre variables  $\lambda(1,1), \lambda(1,2), \lambda(2,1), \lambda(2,2)$ :

$$f(\lambda(1,1), \lambda(1,2), \lambda(2,1), \lambda(2,2)) = c1_s \lambda(1,1)^{n(1,1)} \lambda(1,2)^{n(1,2)} \lambda(2,1)^{n(2,1)} \lambda(2,2)^{n(2,2)}:$$

où  $S$  est le simplexe :

$$\lambda(1,1) \geq 0, \lambda(1,2) \geq 0, \lambda(2,1) \geq 0, \lambda(2,2) \geq 0$$

$$\lambda(1,1) + \lambda(1,2) + \lambda(2,1) + \lambda(2,2) = 1$$

$$\lambda(1,1) \geq \lambda(1,2) \geq \lambda(2,2), \lambda(1,1) \geq \lambda(2,1) \geq \lambda(2,2)$$

### C. Applications

Une fois que la loi conjointe  $f$  est connue, les applications sont immédiates : on peut calculer la probabilité de certains événements portant sur une sous-population (probabilités que certaines stations soient au-dessus de certains seuils), les lois marginales, etc.

Par exemple, dans le cas de deux stations et deux classes, l'espérance de  $\lambda(1,1)$  sera :

$$p(1,1) = \int \lambda(1,1) f(\lambda(1,1), \lambda(1,2), \lambda(2,1), \lambda(2,2)) d\lambda(1,1) d\lambda(1,2) d\lambda(2,1) d\lambda(2,2)$$

Dans le calcul ci-dessus, on constate simplement que le nombre  $n(1,1)$  est augmenté d'une unité.

Cette approche permet en outre d'évaluer la probabilité d'évènements non observés ; pour cela, il suffit de créer une ou plusieurs classes supplémentaires sur chaque station. Le nombre de relevés des évènements non observés (par exemple le triplet 1600, 1500, 2000) sera nul ; ceci ne pose pas de problème pour l'implémentation de la méthode, et le triplet est bien pris en compte dans la construction de la loi conjointe.

On est donc amené, aussi bien pour le calcul de  $f$  que pour celui des espérances des lois marginales, à évaluer l'intégrale multiple d'une fonction de type (1) (qui est un monôme) sur un simplexe de type (2) (comportant de multiples inégalités, correspondant aux comparaisons entre variables).

## D. Calcul de l'intégrale

La méthode que nous présentons s'inspire de celle utilisée par Peter Robinson [Robinson] pour simplifier notre approche dans le cas de la dimension 1 ; elle-même s'inspire des résultats de Luc Devroye [Devroye].

Le calcul de l'intégrale multiple se fait par une méthode de type Monte-Carlo. Posons, pour simplifier les notations :

$$\Lambda = (\lambda(i, j, k))_{i,j,k}$$

Cette notation représente 1000 variables  $\lambda(i, j, k)$  dans le cas de trois stations et dix classes.

On cherche à évaluer  $\int f(\Lambda)d\Lambda$ . On effectue un tirage de  $M$  valeurs de  $\Lambda$  et on remplace l'intégrale par la moyenne calculée sur cet échantillon :

$$\int f(\Lambda)d\Lambda \sim \frac{1}{M} \sum_{m=1}^M f(\Lambda_m)$$

où  $\Lambda_m$  désigne le  $m$ -ème échantillon (un échantillon est ici une valeur attribuée aux 1000 variables qui composent  $\Lambda$ ).

Pour que la méthode de Monte-Carlo soit applicable, il faut que les tirages soient uniformément répartis dans le simplexe (2). Celui-ci comporte un grand nombre de comparaisons relatives à l'ordre.

### 1. Etape 1

On commence par un simplexe plus simple, où tous les points sont ordonnés de manière décroissante.

Notons  $A$  le nombre de variables  $\lambda(i, j, k)$  :  $A = C^3$  (dans le cas de trois stations) ; rappelons que  $C$  désigne le nombre de classes. Introduisons le simplexe :

$$S_0 = \left\{ (x_1, \dots, x_A) ; x_a \geq 0, x_1 \geq x_2 \geq \dots \geq x_A, \sum_{a=1}^A x_a = 1 \right\}$$

Nous allons d'abord tirer un échantillon bien réparti au sein de ce simplexe. Pour cela, selon la méthode de Devroye-Robinson :

- On génère  $A$  nombres  $x_a$  selon une loi uniforme entre 0 et 1.
- On les réarrange de manière décroissante :  $x_1 > x_2 > \dots$
- On considère les  $A$  "spacings" :

$$d_1 = x_{T^2-1}, d_2 = x_{A-2} - x_{A-1}, \dots, d_{A-a+1} = x_{a-1} - x_a, \dots, d_{A-1} = x_1 - x_2, d_A = 1 - x_1$$

On sait qu'ils sont uniformément distribués dans l'ensemble :

$$S_1 = \left\{ x_a \geq 0, \sum_a x_a = 1 \right\}$$

(la transformation qui permet de passer des  $x_a$  aux  $d_a$  a pour Jacobien 1).

- On réarrange ces spacings de manière décroissante ; on les note encore  $d_k$ . On sait que les spacings, une fois réarrangés, sont uniformément distribués dans le simplexe :

$$S_0 = \left\{ (x_1, \dots, x_A); x_a \geq 0, x_1 \geq x_2 \geq \dots \geq x_A, \sum_{a=1}^A x_a = 1 \right\}$$

On obtient ainsi un échantillon noté  $X$  (il est de taille  $A$ ).

## 2. Etape 2

On introduit le simplexe  $S$ . Pour cela, il faut passer de l'ordre total existant dans le simplexe  $S_0$  aux ordres partiels définis par les formules (2).

Les formules (2) se traduisent par le fait que les  $C^3$  nombres  $\lambda(i, j, k)$  sont rangés dans un "Tableau de Young", dont  $\lambda(1,1,1)$  est la base.

En dimension 2, un tableau de Young est une matrice, de taille  $C \times C$ , à l'intérieur de laquelle on a rangé les nombres  $1, 2, \dots, C^2$  de telle manière que, sur chaque ligne et chaque colonne, les nombres soient croissants. Bien entendu, 1 se trouve en haut à gauche et  $C^2$  se trouve en bas à droite. Voici un exemple pour  $C = 3$ :

1	2	6
3	5	8
4	7	9

Si l'on range des nombres décroissants selon un tableau de Young (le  $k$ -ème nombre étant mis dans la case numérotée  $k$ ) on obtient un tableau dont les lignes et les colonnes sont décroissantes.

En dimension 3, la même construction donne un tableau dont les nombres sont décroissants selon chaque axe, ce qui rend compte des formules (2).

Le nombre de Tableaux de Young est très élevé ; il est donné par une formule explicite, appelée "hook-length formula". De manière algorithmique, il n'est pas possible de les générer tous. Nous procéderons donc, là encore, à des tirages aléatoires.

A chaque fois qu'un échantillon  $X$  est généré, selon la méthode décrite à l'étape 1, on génère aussi un Tableau de Young  $T$ , et on range l'échantillon  $X$  dans le tableau, selon l'ordre de ce tableau. Il faut évidemment s'assurer que les Tableaux de Young sont générés selon une loi uniforme (ne pas privilégier certains tableaux). Comme les nombres générés au sein de l'échantillon  $X$  sont des nombres réels décroissants, les cas d'égalité sont de mesure nulle ; autrement dit on peut supposer que les nombre de  $X$  sont strictement décroissants, et il n'y a pas d'ambiguïté à propos de la manière de les ranger.

### 3. Génération de Tableaux de Young - algorithmes

Sur un plan algorithmique, cette génération doit être faite très rapidement, parce que nous aurons besoin d'un tableau pour chaque échantillon  $X$  décrit ci-dessus. On ne peut donc pas se permettre, à chaque étape, d'explorer l'ensemble des cases du tableau pour décider de l'étape suivante. La méthode que nous présentons ici est donc purement locale, comme on va le voir : elle ne repose que sur l'examen des cases voisines (en l'occurrence 3 en dimension 3) d'une case donnée.

Dans ce paragraphe, nous adoptons des notations un peu différentes de celles des paragraphes précédents. Nous notons  $K = C^3$  le nombre total de cases du tableau.

Rappelons que nous devons ranger les nombres  $1, 2, \dots, K$  dans le tableau cubique, de telle sorte que les nombres soient croissants le long de chaque axe. Le nombre 1 est évidemment mis dans la case de coordonnées  $(1, 1, 1)$  ; pour le nombre 2, il y a trois choix possibles, à savoir  $(2, 1, 1), (1, 2, 1), (1, 1, 2)$ , et ainsi de suite pour les suivants.

Admettons que nous ayons rempli le tableau jusqu'à l'entier  $k$  ; nous voulons savoir où mettre l'entier  $k + 1$ . Nous dirons que nous sommes à l'étape  $k$ .

Nous savons où nous avons rangé les  $k$  premiers entiers. Nous allons coder cette information sous deux formes : elles sont mathématiquement équivalentes, mais chacune a son intérêt algorithmique. Les algorithmes sont donnés ici sous VBA pour Excel (sous Matlab par la suite).

Tout d'abord, nous créons un array :

```
dim cse(1 to C, 1 to C, 1 to C) as integer
```

dans lequel nous mettons le numéro de la case, lorsque celui-ci est attribué. On a par exemple  $cse(1, 1, 1) = 1$  et si on a choisi  $(2, 1, 1)$  pour mettre le nombre 2, on aura  $cse(2, 1, 1) = 2$ . L'array  $cse$  est initialisé à 0 (ou plus exactement il n'est pas initialisé du tout), ce qui fait que seules les cases attribuées ont un  $cse > 0$ .

Ensuite, nous créons les arrays

```
dim coord_cse(1 to K, 1 to V) as integer
```

dans lesquels nous mettons les coordonnées de la case choisie à chaque étape.  $coord\_cse(k, v)$  désigne la  $v$ -ème coordonnée ( $v = 1, 2, 3$ ) de la  $k$ -ème case.

Par exemple, si on a choisi  $(2, 1, 1)$  pour mettre le nombre 2, on aura :

```
coord_cse(2, 1) = 2, coord_cse(2, 2) = 1, coord_cse(2, 3) = 1.
```

Ces trois arrays constituent en quelque sorte l'application inverse de "cse".

Lorsque nous sommes à l'étape  $k$ , nous connaissons donc :

- Tous les  $cse(i_1, i_2, i_3)$  (seuls  $k$  d'entre eux sont  $> 0$ )
- Les  $coord\_cse(j, 1), coord\_cse(j, 2), coord\_cse(j, 3)$ , pour  $j = 1, \dots, k$ .

Nous disposons aussi, à l'étape  $k$ , d'une liste de "case candidates", c'est-à-dire susceptibles de recevoir l'entier  $k+1$ . Une case, de coordonnées  $(u, v, w)$ , est candidate si et seulement si ses précédents sont occupés, c'est-à-dire si les trois cases  $(u-1, v, w), (u, v-1, w), (u, v, w-1)$  sont occupées. Lorsque l'un ou plusieurs de ces précédents n'existent pas ( $u=1, v=1, w=1$ ), c'est-à-dire lorsque la case à tester est située sur le bord, la condition correspondante est omise (en d'autres termes, on considère le bord comme toujours occupé).

Voici un exemple en dimension 2 ;

1	4	x	
2	5		
3	x		
x			

Nous avons rempli le tableau jusqu'à l'entier 5. Les cases candidates sont au nombre de 3 ("x" dans le tableau), dont 2 sont près du bord. Une case ne peut être candidate que si ses deux antécédents (à gauche et en haut) sont occupés. Si la case est située près du bord, la condition correspondante est automatiquement satisfaite.

Admettons donc que, à l'étape  $k$ , nous ayons une liste de cases candidates : soit  $N_c$  le nombre de cases candidates à l'étape  $k$ . Il est impératif de disposer d'une liste (c'est-à-dire d'un ensemble numéroté de 1 à  $N_c$ ) et non simplement d'un ensemble ordonné. En effet, nous allons tirer selon une loi uniforme un nombre entier entre 1 et  $N_c$ .

A l'étape  $k$ , les  $N_c$  cases candidates sont données par leurs coordonnées :

```
dim cand(1 to K, 1 to V) as integer
```

(avec ici  $V = 3$ , nombre de villes)

$cand(n,1)$  désigne la première coordonnée de la  $n$ -ème case ( $n=1, \dots, N_c$ ), et de même pour  $cand(n,2), cand(n,3)$ . Les arrays sont définis une fois pour toute jusqu'à  $K$  (nombre total de cases), mais à l'étape  $k$  ils ne sont remplis que jusqu'à  $N_c$ ; les suivants sont laissés à 0.

Donc, à l'étape  $k$ , nous disposons de la liste des cases candidates et de leur nombre  $N_c$ . Nous tirons avec loi uniforme un nombre entre 1 et  $N_c$ ; soit  $n_0$  le nombre ainsi choisi. La case ainsi déterminée va recevoir le nombre  $k+1$ . Ceci se traduit par :

```
cse(cand(n_0,1),cand(n_0,2),cand(n_0,3)) = k + 1
coord_cse(k+1,v) = cand(n_0,v), v = 1, ..., 3
```

A ce stade, la liste des cases occupées a donc été incrémentée d'une unité. Reste à modifier la liste des cases candidates, tout en conservant une liste. Or l'élément  $n_0$  a disparu, puisqu'il vient d'être choisi. Mais il a pu générer d'autres cases candidates : au plus trois, qui sont ses descendants, s'ils existent.

Les "descendants" d'une case  $(u, v, w)$  sont les trois cases  $(u+1, v, w), (u, v+1, w), (u, v, w+1)$ . Pour qu'ils figurent dans la liste des cases candidates (nous dirons alors qu'ils sont "éligibles"), il faut simplement que leurs antécédents soient des cases occupées.

Une case ayant été choisie, nous avons donc les opérations suivantes à effectuer :

- Déterminer ses descendants (qui peuvent être au nombre de 0, 1, 2, 3) ;
- Pour chacun de ces descendants, déterminer s'il est éligible ou non, c'est-à-dire déterminer si les antécédents de ces descendants sont occupés (deux opérations à chaque fois) ;
- Reprendre la liste des cases candidates et y insérer les nouvelles cases éligibles, en faisant le moins de modifications possible à la liste (il faut impérativement éviter de tout renumérotter). Nous montrons que ceci requiert au plus trois opérations.

En d'autres termes, l'ensemble du calcul de la nouvelle liste candidate requiert au plus 9 opérations.

Voici d'abord une description logique de l'algorithme :

Si la case choisie n'a aucun descendant, le dernier élément de la liste est mis à la place de l'entier  $n_0$  et la liste est raccourcie d'une unité ( $Nc = Nc - 1$ ). Ceci se produit dans l'exemple suivant :

1	2	4	x5
3	x		
x			

Nous avons choisi 1,2,3,4, et les cases candidates sont alors désignées par x. Nous choisissons de mettre 5 à droite de la première ligne. La case numérotée x5 n'a aucun descendant. Nous n'avons plus, alors que deux cases candidates (les deux x précédents).

Si la case retenue a exactement un descendant, celui-ci est mis à sa place et le nombre d'éléments de la liste est conservé.

Si la case retenue a deux descendants, le premier est mis à sa place et le second est mis en dernier ;  $Nc$  est incrémenté d'une unité.

Si la case retenue a trois descendants, le premier est mis à sa place et les deux derniers sont mis en dernier ;  $Nc$  est incrémenté de deux unités.

#### 4. Description fonctionnelle de l'algorithme :

Notons  $(u, v, w)$  les coordonnées de la cellule choisie à l'étape  $k$ . Nous devons tester les cellules  $C_1 = (u+1, v, w), C_2 = (u, v+1, w), C_3 = (u, v, w+1)$  si elles existent, c'est-à-dire si la cellule choisie n'est pas proche du bord ( $u < C, v < C, w < C$ ).

Test d'éligibilité de  $C_1$  : si elle existe et si ses deux antécédents,  $(u+1, v-1, w), (u+1, v, w-1)$  sont occupés. Une case est occupée si  $cse(x, y, z) > 0$ .

Nous avons donc le test suivant pour  $C_1$  :

```

if (u < C) then
If (v = 1 Or (v > 1 And cse(u + 1, v - 1, w) > 0)) And (w = 1 Or (w > 1 And cse(u + 1, v, w - 1) > 0)) Then
test1 = 1
end if 'if v=1 etc
end if 'if u<C

```

Pour C2

```

if (v < C) then
If (u = 1 Or (u > 1 And cse(u - 1, v + 1, w) > 0)) And (w = 1 Or (w > 1 And cse(u, v + 1, w - 1) > 0)) Then
test2 = 1
end if 'if u=1 etc
end if 'if v<C

```

Pour C3

```

if (w < C) then
if (u=1 or (u>1 and cse(u-1,v,w+1)>0) and (v=1 or (v>1 and cse(u, v-1, w+1)>0) ) ) then
test3 = 1
end if 'if u=1 etc.
end if 'if w<C

```

Ensuite, on procède de la manière suivante :

```

if test1= 1 (C1 éligible)
on la met dans la liste à la place de cand(n0,v)
cand(n0,v) = C1(v), v=1,2,3
if test2=1 (C2 éligible aussi)
Nc=Nc+1
cand(Nc,v)=C2(v), v=1,2,3 'on rajoute C2 au bout de la liste des éligibles
end if 'test2=1
if test3=1
Nc=Nc+1
cand(Nc,v)=C3(v), v=1,2,3
'on rajoute C3 au bout de la liste des éligibles, que C2 le soit ou non
end if 'test3=1
end if 'if test1=1

if test1=0 'C1 pas éligible
if test2=1
cand(n0,v)=C2(v), v=1,2,3 'on met C2 à la place de la case choisie
endif 'test2=1

if test2=0 'C2 pas éligible non plus
if test 3 = 1 'C3 éligible
cand(n0,v)=C3(v), v=1,2,3
end if 'test3=1

if test3 = 0 'aucune case n'est éligible
cand(n0,v) = cand(Nc, v), v=1,2,3
'on met la dernière case à la place de la case choisie
Nc=Nc-1 'le nombre de cases éligibles décroît d'une unité

```

```
endi if 'test3=0  
end if 'test1=0
```

Nombre total d'opérations à l'étape  $k$  :

Tirer un nombre aléatoire entre 1 et  $N_c$  : 1 opération

Reporter la case choisie dans les cases occupées : 4 opérations scalaires

Déterminer les descendants de la case choisie et tester s'ils sont éligibles : au plus 9 opérations

Reconstituer la liste des cases candidates : au plus 3 opérations

Total : au plus 17 opérations à chaque étape. Le nombre d'étapes est égal au nombre de cases du tableau, soit  $C^3$  ; il faut donc au plus  $17 C^3$  opérations pour la génération de chaque échantillon, avec le tableau de Young associé.

## VI. Extension de la loi conjointe : mise en œuvre pratique

Nous présentons ici les étapes principales de la construction de la loi conjointe, pour les phénomènes extrêmes, en suivant la théorie développée plus haut. Mais dans un premier temps, mentionnons les difficultés rencontrées.

### A. Les difficultés

La construction d'une loi de probabilité standard ne dépend que des données, et non de la localisation des stations. Dans le cas d'une seule station, pour les phénomènes extrêmes, il en allait de même : les données fournies représentent toute l'information disponible.

Mais les choses seront différentes pour la loi conjointe : il faudra étudier les particularités régionales, comme nous le verrons. La taille de l'historique des différentes stations entre aussi en considération.

Dans un premier temps, nous cherchons à observer le changement de la loi de probabilité en fonction de la taille de l'historique étudié. Pour le voir, comparons une situation où nous travaillons sur  $N$  années, et une autre où nous travaillons sur  $M+N$  années.

#### 1. Cas de la station de Brest

Considérons d'abord le cas de Brest ; l'historique est le plus long. Prenons les valeurs minimales et maximales de la hauteur observée et divisons cet intervalle en 20 intervalles égaux. Ainsi, nous construisons les lois de probabilité standards pour deux historiques :

- période janvier 1846 à mai 1857 (11,5 ans) ;
- période complète de 151 ans.

La figure ci-dessous représente les deux histogrammes obtenus.

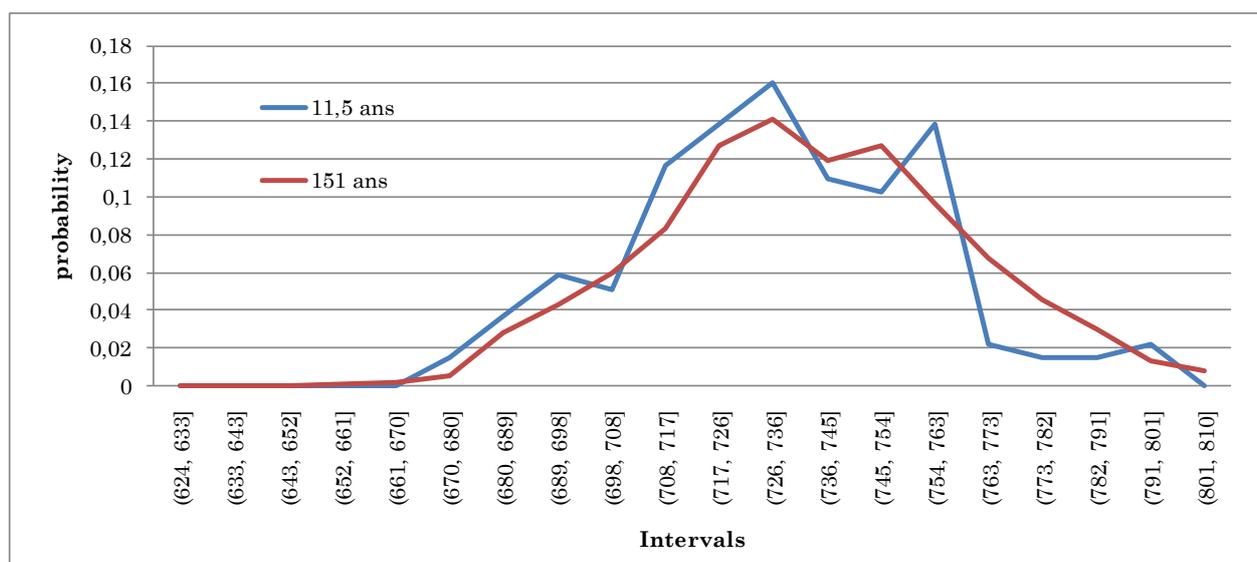


Figure 22 : Lois de probabilité standards de la station de Brest

Il y a des différences significatives pour les valeurs fortes (740-780], mais guère pour les valeurs les plus fortes (780-810]. Pour la station de Brest, nous ne remarquons pas de phénomènes extrêmes. C'est pourquoi, les lois de probabilité sont presque identiques pour les valeurs les plus fortes.

Sur la période complète, la station de Brest n'a jamais enregistré de valeurs particulièrement élevées, si bien qu'il est difficile de décider jusqu'à quelle limite la probabilité doit être recherchée. Cependant, il n'existe pas de raison d'exclure un orage particulièrement violent à cet endroit.

## 2. Cas de la station de La Rochelle

La seconde station est La Rochelle. L'historique total inclut 70 années d'observations, sur lesquelles seulement 28 sont disponibles. La première période est de mars 1967 à février 1980 (11,2 années) et la seconde est de juin 1946 à juin 2010 (28 années). Nous obtenons les histogrammes suivants :

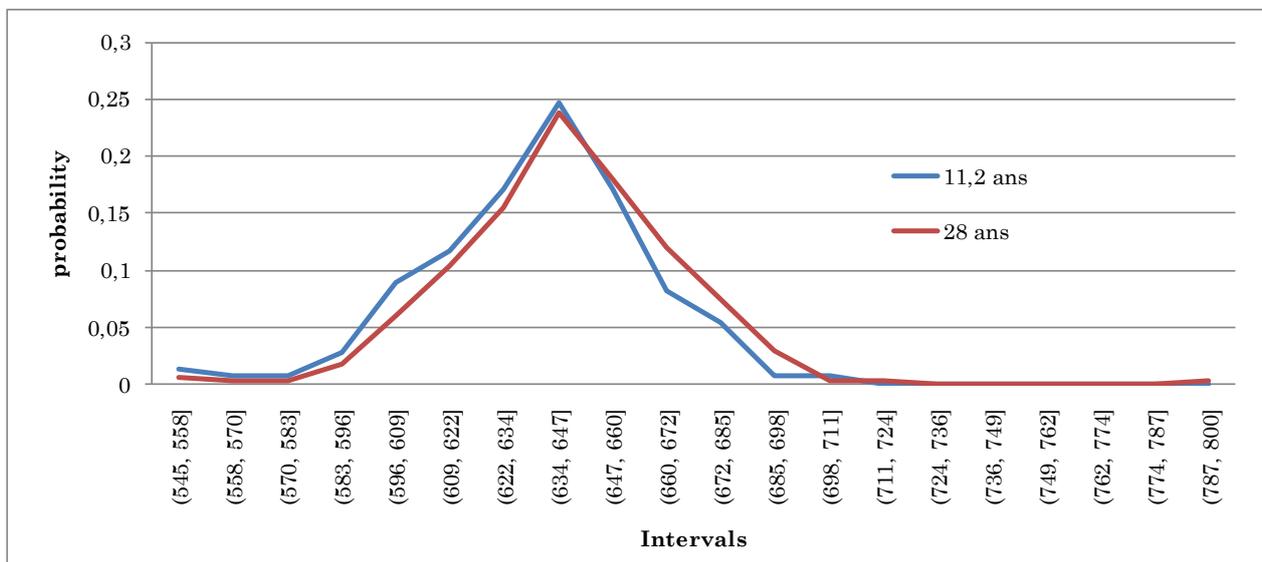


Figure 23 : Loïs de probabilité standards de la station de la Rochelle

L'histogramme complet prend en compte la tempête Xynthia en 2010, qui n'apparaît pas sur le premier. Cette tempête implique un décalage vers la droite de la loi de probabilité. L'observation de 11 premières années n'est donc pas suffisante pour décrire la loi de probabilité déterminée sur un historique plus important.

## 3. Conclusion

Dans les deux cas, les courbes deviennent moins concentrées lorsque l'intervalle de temps augmente. Cependant, pour les phénomènes rares, nous constatons sur ces exemples que l'évaluation des valeurs au cours du temps n'a pas le même comportement pour les deux stations. L'application de la méthode d'extension de la loi conjointe dépendra des caractéristiques de l'historique de chaque station.

Notre approche générale sera la suivante :

- Si une station a enregistré des valeurs particulièrement élevées, nous l'utilisons comme station de référence pour toute la région (pour les trois ou quatre stations proches), en définissant ainsi une limite absolue pour la hauteur d'eau. Il est également possible d'utiliser la connaissance d'expert pour décider si certaines stations sont protégées ou non.
- Si une station n'a pas enregistré de valeurs élevées, nous utilisons la station de référence la plus proche.

On constate donc qu'il est nécessaire d'introduire une certaine connaissance du terrain, ne serait-ce que le degré de proximité géographique entre stations. La seule utilisation des données brutes ne suffit pas.

## *B. Construction de la loi de probabilité standard étendue.*

### **1. Application à la station d'Ajaccio**

La hauteur maximale mensuelle des marées est en moyenne de 64.5 cm. Une forte tempête s'est produite le 19 février 2010, et le niveau de l'eau a atteint la cote 125 cm ; lors de cet événement, la mer a été qualifiée de "très forte".

Pour le groupe des stations de la Méditerranée, cette station sera utilisée comme référence parce que c'est celle où la différence est la plus forte entre la valeur moyenne et le maximum.

Nous définissons d'abord un phénomène extrême comme ayant une probabilité dans les 10 % les plus élevés. La prise en compte de ce seuil a plusieurs défauts :

- Habituellement, il y a peu de données disponibles, et de ce fait la loi sera irrégulière ;
- La valeur du quantile 90 % est très sensible au changement d'historique utilisé pour la construction, surtout lorsqu'on s'intéresse à la loi conjointe (beaucoup de cellules seront vides). L'exemple de La Rochelle montre bien que le quantile à 90 % se déplace significativement en fonction de l'augmentation de la taille de l'historique.

Nous décidons de fixer un autre seuil, à savoir l'espérance (valeur moyenne) ou la médiane. Ce type de seuil est plus stable ; on constate que les valeurs sont effectivement décroissantes à partir de ce seuil (ce qui n'est pas évident en théorie). Dans notre exemple, nous prendrons donc toutes les données supérieures à la moyenne : le domaine étudié est stable en fonction de l'évolution de l'historique. Cette stabilité est importante lorsque la loi conjointe est étendue pour plusieurs stations.

Le choix de la discrétisation est fait de manière empirique. Nous prenons d'abord un intervalle de valeur moyenne (environ 13 cm). Nous considérons l'intervalle ]64,5 cm-133 cm], que nous divisons en cinq classes de même largeur en excluant la borne gauche et en incluant la borne droite.

## Construction des lois de probabilité standard et étendue aux phénomènes extrêmes

Nous prenons les notations suivantes :

- $N$  est le nombre total d'observations disponibles :  $N = 263$
- $K$  est le nombre d'intervalles (classes) considérés :  $K = 5$
- $n_k$  est le nombre des hauteurs observées dans chaque intervalle  $k$  de sorte que
$$\sum_{k=1}^K n_k = 135$$
- $p_k$  est la probabilité caractérisant chaque intervalle que nous cherchons à déterminer.

D'après la théorie présentée au chapitre V, la probabilité (qui est également une variable aléatoire) d'être dans un intervalle  $k$  est égale à la densité de probabilité suivante :

$$p_k = f_{n_k, N}(\lambda) = c \lambda^{n_k} (1 - \lambda)^{N - n_k}$$

où  $\lambda$  est le "taux de risque" associé à chaque intervalle et inconnu pour le moment. Comme nous pouvons l'observer (et c'est une remarque importante), cette probabilité n'est jamais nulle, même si  $n_k = 0$ .

La loi de probabilité de l'ensemble  $(p_1, \dots, p_K)$  est donnée par la formule suivante :

$$f(\lambda_1, \dots, \lambda_K) = c \mathbb{1}_S \lambda_1, \dots, \lambda_K \lambda_1^{n_1} \times \dots \times \lambda_K^{n_K}$$

La loi marginale de chaque  $p_k$  a comme densité :

$$f_k(\lambda) = \int \dots \int_S f(\lambda_1, \dots, \lambda_{k-1}, \lambda, \lambda_{k+1}, \dots, \lambda_K) d\lambda_1 \dots d\lambda_{k-1} d\lambda_{k+1} \dots d\lambda_K$$

Et finalement, la valeur assignée à chaque  $p_k$  sera l'espérance :

$$\overline{p_k} = \int_0^1 \lambda_k f(\lambda_1, \dots, \lambda_K) d\lambda_1 \dots d\lambda_K$$

Pour une meilleure compréhension de l'ensemble des calculs, nous décrivons en détail chaque étape de la construction.

*Etape 1: discrétisation en classes*

Nous discrétisons en  $k$  classes et nous déterminons chaque  $n_k$  :

Classes (Ajaccio), cm				$n_k$
Class 1	]65	-	78]	113
Class 2	]78	-	92]	19
Class 3	]92	-	105]	2
Class 4	]105	-	119]	0
Class 5	]119	-	133]	1

Tableau 31 : Discrétisation en 5 classes

*Etape 2: Initialisation du taux de risque*

Nous devons déterminer le taux de risque  $\lambda_k$  de chaque intervalle. Le fait de traiter les phénomènes extrêmes permet de considérer les taux de risque décroissants tel que :

$$\lambda_1(class1) \geq \lambda_2(class2) \geq \dots \geq \lambda_5(class5) , \lambda_k \geq 0 \text{ and } \sum_{k=1}^5 \lambda_k = 1$$

- a) On génère  $K = 4$  nombres  $x_k$  selon une loi uniforme entre 0 et 1 et on les réarrange de manière décroissante :  $x_1 \geq x_2 \geq \dots \geq x_4$
- b) On considère les 5 "spacings"  $d_k$  :

$$d_1 = x_1, d_2 = x_2 - x_1, \dots, d_{K-1} = x_{K-1} - x_{K-2}, d_K = 1 - x_{K-1}$$

et on réarrange ces « spacings » de manière décroissante.

- c) Au final, nous associons chaque  $d_k$  à chaque taux de risque  $\lambda_k$ , tel que :

Classes (Ajaccio), cm				$\lambda_k$
Class 1	]65	-	78]	0,477
Class 2	]78	-	92]	0,390
Class 3	]92	-	105]	0,106
Class 4	]105	-	119]	0,025
Class 5	]119	-	133]	0,003

Tableau 32 : Initialisation des taux de risque

*Etape 3 : Calcul de la loi de probabilité de l'ensemble  $(p_1, \dots, p_K)$  :*

La loi de probabilité de l'ensemble  $(p_1, \dots, p_K)$  est donnée par la formule suivante :

$$f(\lambda_1, \dots, \lambda_K) = c \mathbb{1}_{S(\lambda_1, \dots, \lambda_K)} \lambda_1^{n_1} \dots \lambda_K^{n_K}$$

où  $c$  est le coefficient de normalisation qui doit être calculé et  $S = \mathbb{1}_{S(\lambda_1, \dots, \lambda_K)}$  est un indicateur qui représente toutes les combinaisons possibles  $\lambda_k$  que nous appelons "simplexe de travail".

Sur l'ensemble de l'espace considéré, l'intégrale multiple de la fonction  $f$  est égale à 1.

$$\int_S \dots \int f(\lambda_1, \dots, \lambda_K) d\lambda_1 \dots d\lambda_K = 1$$

Nous utilisons la méthode de Monte Carlo pour calculer l'intégrale :

$$\int_S \dots \int f(\lambda_1, \dots, \lambda_K) d\lambda_1 \dots d\lambda_K \approx c \cdot \frac{1}{M} \sum_{m=1}^M \lambda_{m,1}^{n_1} \times \dots \times \lambda_{m,K}^{n_K} = 1$$

où pour chaque  $m$ , nous générons une nouvelle série de  $\lambda_k$ . Le nombre  $M$  doit être raisonnablement important (1 000 000 ou 5 000 000) afin d'étudier de manière précise l'ensemble des séries de  $\lambda_k$  possibles. Ainsi, nous obtenons le coefficient de normalisation  $c$ .

*Etape 4 : Estimation de la loi marginale :*

L'estimation de la loi marginale de chaque  $p_k$  est ainsi obtenue de la manière suivante :

$$\overline{p_k} = \int_S \dots \int \lambda_k f(\lambda_1, \dots, \lambda_K) d\lambda_1 \dots d\lambda_K = c \cdot \frac{1}{M} \sum_{m=1}^M \lambda_{m,k} \times (\lambda_{m,1}^{n_1} \times \dots \times \lambda_{m,K}^{n_K})$$

Les probabilités  $p_k$  obtenues par les deux méthodes sont comparées dans le tableau ci-dessous :

Classes (Ajaccio), cm			Méthode standard	Méthode étendue aux phénomènes extrêmes
Class 1	]65	- 78]	0,430	0,419
Class 2	]78	- 92]	0,072	0,073
Class 3	]92	- 105]	0,008	0,013
Class 4	]105	- 119]	0	0,006
Class 5	]119	- 133]	0,004	0,003

Tableau 33 : Comparaison de la loi standard et de la loi étendue aux phénomènes extrêmes

Il faut bien remarquer que la somme des probabilités n'est pas égale à 1 : en effet, dans notre exemple, cette somme correspond à la probabilité d'être supérieur à l'espérance. De cette manière, la probabilité de chaque classe est directement obtenue.

Voici l'histogramme obtenu :

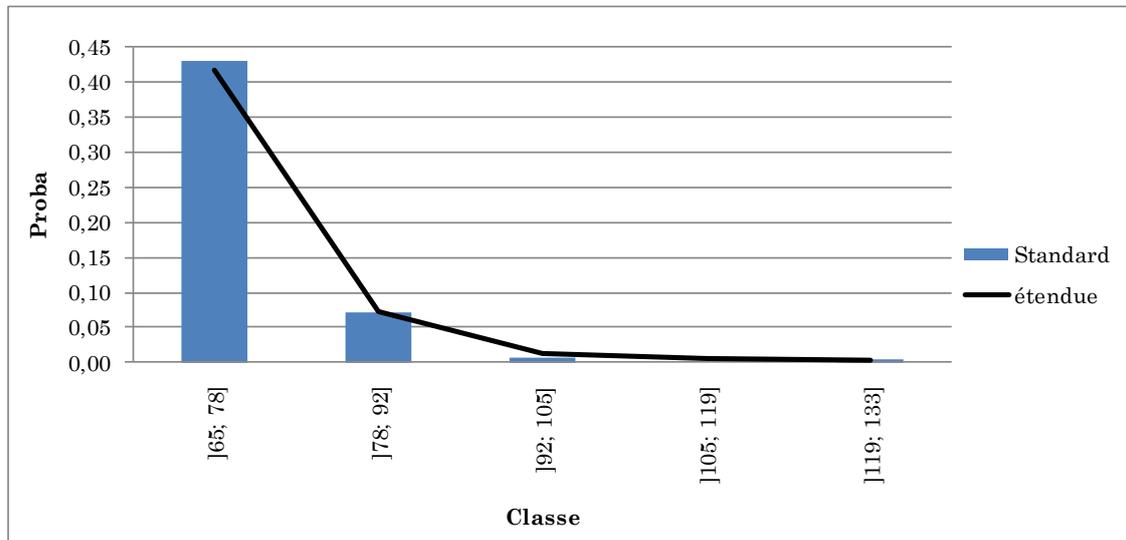


Figure 24 : Lois de probabilité standard et étendue de la station d'Ajaccio

Nous observons que les deux méthodes donnent des résultats similaires. Cependant, la méthode étendue aux phénomènes extrêmes est basée sur l'hypothèse de décroissance de probabilité des événements de plus en plus extrêmes. Les résultats de cette méthode sont plus robustes car un très grand nombre de mesures peuvent être simulées. Ainsi la probabilité de la classe ]105 cm; 119 cm) n'est plus nulle, elle est estimée à 0,006.

### Ajout des classes supérieures

Le phénomène extrême relatif à la station d'Ajaccio (niveau 125 cm) sert de référence. Nous fixons la probabilité associée à l'intervalle ]119-131], égale à 0,004 comme base de calcul ; nous ajoutons une classe après celle-là et nous utilisons la méthode de prolongement décrite dans le chapitre V pour calculer les probabilités de chaque classe. Si, après ce calcul, la valeur attribuée à l'intervalle ]119-133] n'est pas approximativement 0,004, nous rajoutons encore une classe supérieure.

Dans le cas présent, nous constatons que le nombre de classes vides à rajouter est de 2, ce qui nous fait au total 7 classes. Nous comparons la loi standard et la loi calculée dans le graphe suivant :

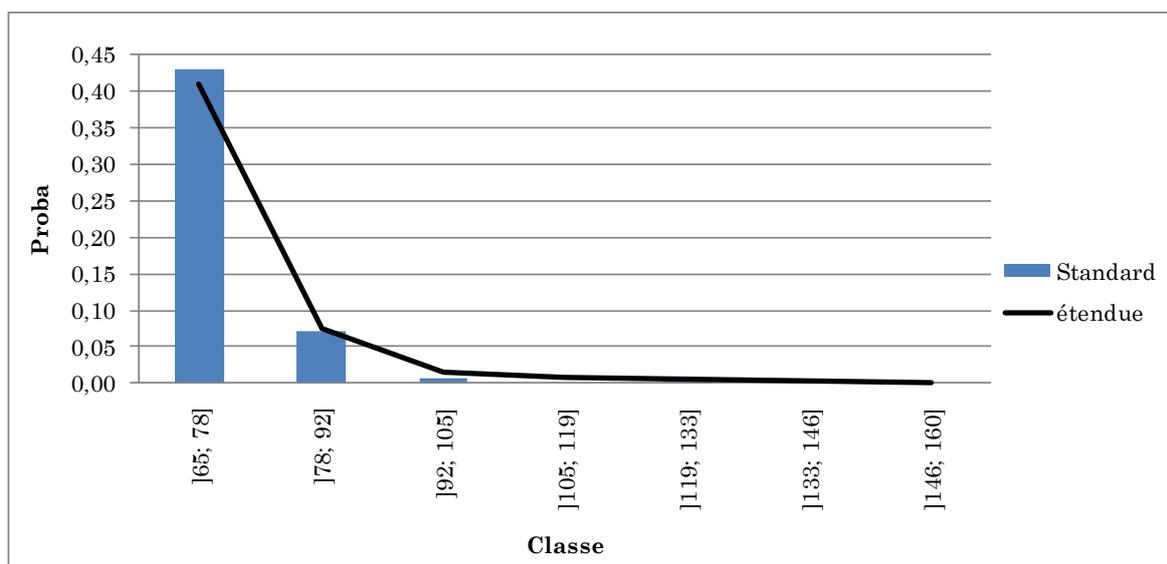


Figure 25 : Lois de probabilité standard et étendue aux phénomènes extrêmes de la station d'Ajaccio

Le tableau ci-dessous présente les valeurs numériques :

	Classe	Standard	étendue
1	]65; 78]	0,430	0,411
2	]78; 92]	0,072	0,074
3	]92; 105]	0,008	0,014
4	]105; 119]	0	0,007
5	]119; 133]	0,004	0,004
6	]133; 146]	0	0,0023
7	]146; 160]	0	0,0009

Tableau 34 : Probabilité des classes des lois standard et étendue pour la station d'Ajaccio

Par exemple, à partir des résultats obtenus pour la loi de probabilité étendue, nous pouvons conclure que la période de retour de la classe la plus élevée (146, 160] est égale à 93 ans, soit environ un siècle.

Pour calculer la loi de probabilité des autres stations du même groupe, nous pouvons procéder de la manière suivante :

- Soit nous procédons comme pour la station d'Ajaccio en ajoutant le nombre de classes nécessaires en fonction de la dernière classe non nulle de la loi de probabilité standard ;
- Soit nous utilisons l'écart relatif entre l'espérance et le maximum observé dans la station d'Ajaccio pour déterminer le nombre de classes à ajouter ;
- Soit nous fixons le maximum absolu en fonction d'un avis d'expert.

## 2. Application à la station de Monaco

La hauteur maximale mensuelle des marées est en moyenne de 69 cm et la hauteur maximale enregistrée est de 105 cm. Les événements extrêmement rares n'ont jamais été observés dans cette région. En fonction de l'allure de la probabilité standard, nous choisissons de ne pas utiliser la station d'Ajaccio comme référence. L'extension est donc paramétrée par la dernière classe non nulle de la loi de probabilité standard.

Nous fixons la taille de classe égale à 13,5 cm et nous calculons la probabilité standard. La probabilité de la dernière classe non nulle est 0,019.

En ajoutant trois classes supérieures, nous obtenons une loi de probabilité étendue cohérente avec la loi standard :

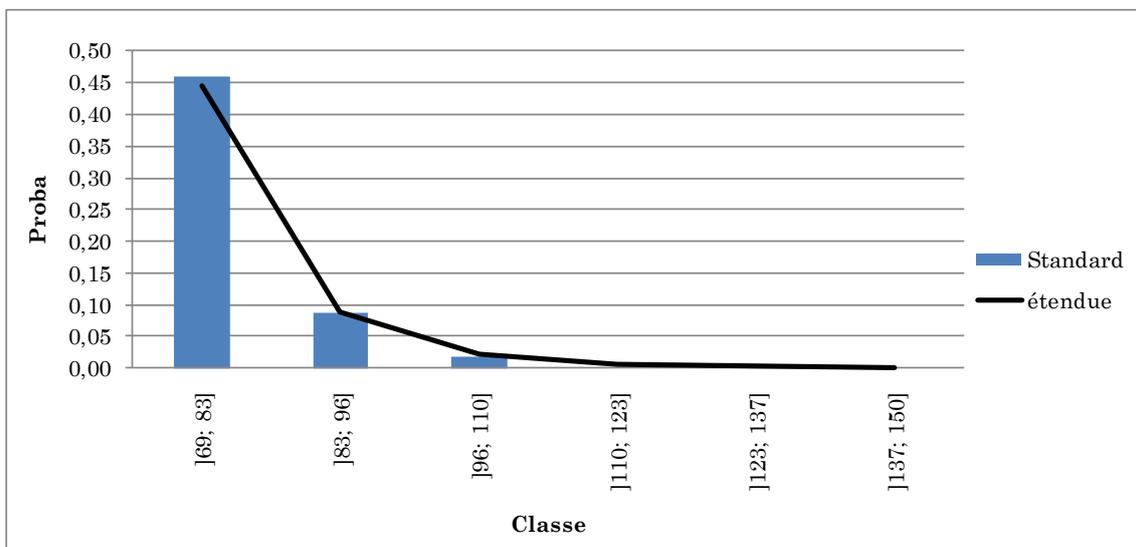


Figure 26 : Lois de probabilité standard et étendue aux phénomènes extrêmes de la station de Monaco

Le tableau ci-dessous présente les valeurs numériques :

	Classe Monac	Standard	étendue
1	]69; 83]	0,460	0,446
2	]83; 96]	0,087	0,087
3	]96; 110]	0,019	0,023
4	]110; 123]	0	0,0065
5	]123; 137]	0	0,0030
6	]137; 150]	0	0,0013

Tableau 35 : Probabilité des classes des lois standard et étendue pour la station de Monaco

### 3. Application à la station de Marseille

La hauteur maximale mensuelle des marées est en moyenne de 71 cm et la hauteur maximale enregistrée est de 137 cm.

Nous fixons la taille de classe égale à 15,5 cm et nous calculons la probabilité standard. La probabilité de la dernière classe non nulle est 0,011.

En ajoutant deux classes supérieures, nous obtenons une loi de probabilité étendue cohérente avec la loi standard :

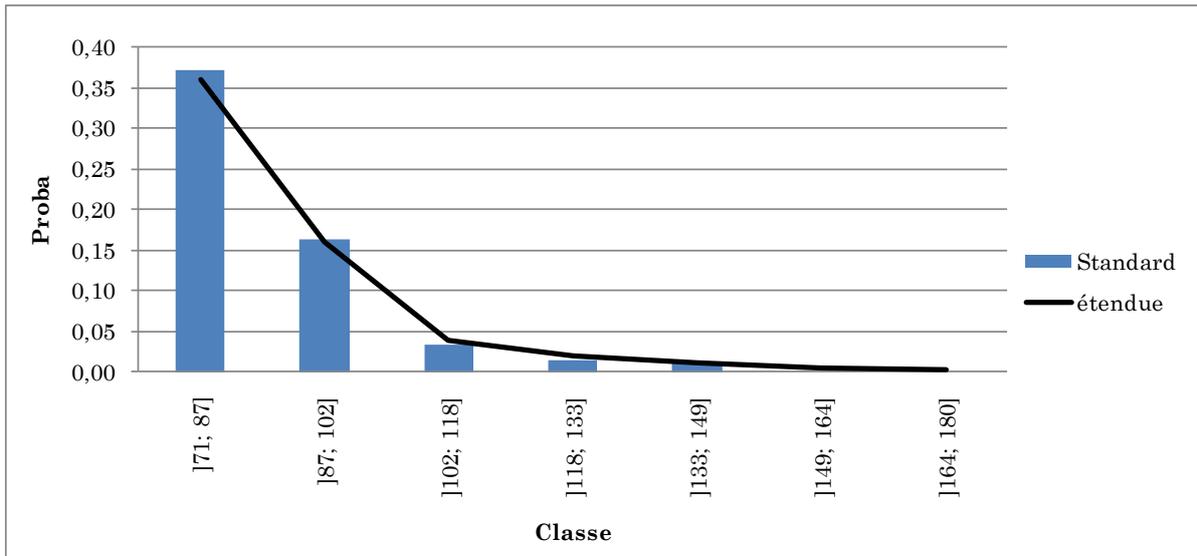


Figure 27 : Lois de probabilité standard et étendue aux phénomènes extrêmes de la station de Marseille

Le tableau ci-dessous présente les valeurs numériques :

	Classe Marse	Standard	étendue
1	]71; 87]	0,373	0,360
2	]87; 102]	0,163	0,160
3	]102; 118]	0,034	0,038
4	]118; 133]	0,015	0,020
5	]133; 149]	0,011	0,012
6	]149; 164]	0	0,0044
7	]164; 180]	0	0,0016

Tableau 36 : Probabilité des classes des lois standard et étendue pour la station de Marseille

### C. Construction de la loi conjointe étendue pour deux stations

Nous commençons par analyser la loi conjointe de deux stations. Nous prenons comme exemple les stations d'Ajaccio et Monaco qui sont dépendantes l'une de l'autre.

Les deux historiques disponibles sont représentés dans la figure ci-dessous :

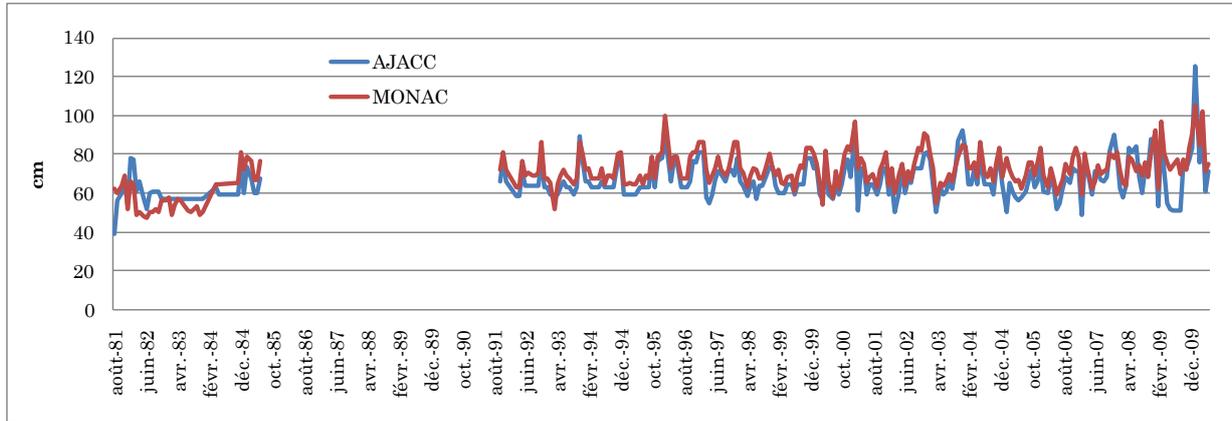


Figure 28 : Historique des stations d'Ajaccio et Monaco

En considérant les classes définies précédemment pour chacune des stations, nous construisons la loi conjointe standard de ces deux stations. La figure ci-dessous représente la loi conjointe affichée en trois dimensions.

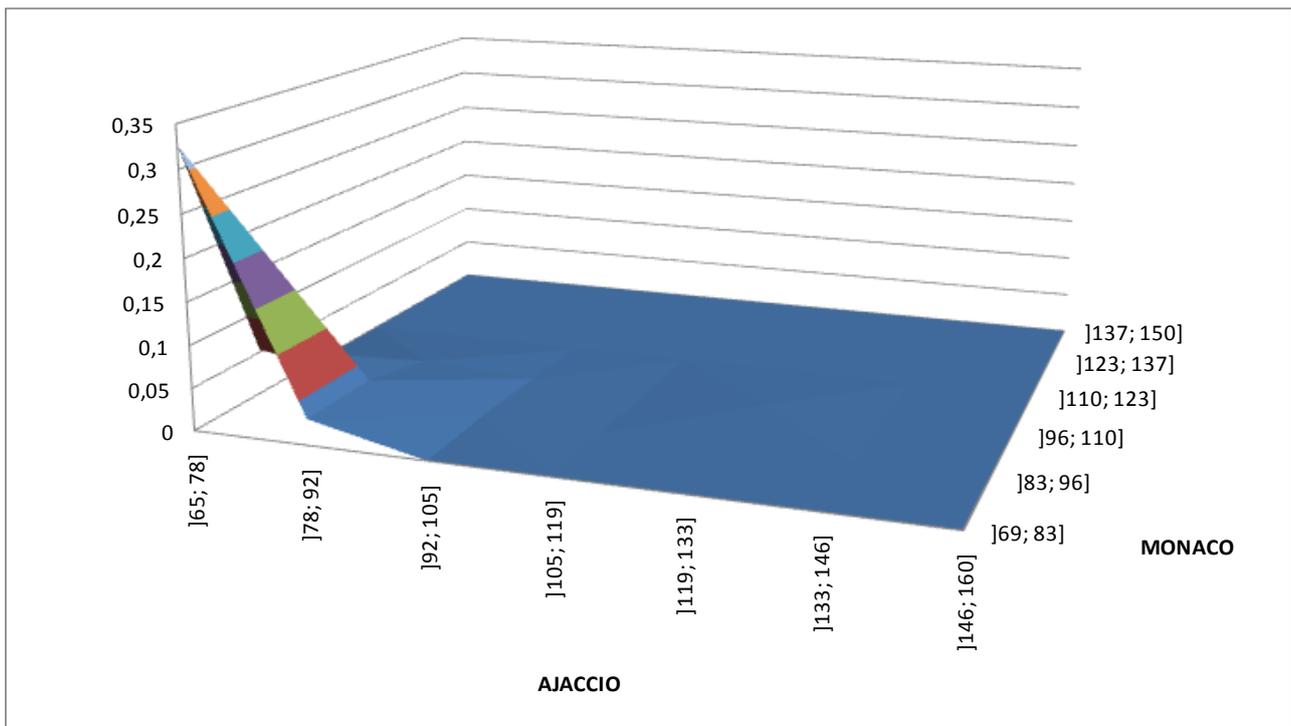


Figure 29 : Loi conjointe standard des stations d'Ajaccio et Monaco

La loi conjointe standard peut être affichée sous la forme d'un tableau :

Ajacc\Monac	]69; 83]	]83; 96]	]96; 110]	]110; 123]	]123; 137]	]137; 150]	somme
]65; 78]	0,3269962	0,0532319	0	0	0	0	0,3802
]78; 92]	0,0304183	0,0304183	0,0114068	0	0	0	0,0722
]92; 105]	0	0,0038023	0,0038023	0	0	0	0,0076
]105; 119]	0	0	0	0	0	0	0
]119; 133]	0	0	0,0038023	0	0	0	0,0038
]133; 146]	0	0	0	0	0	0	0
]146; 160]	0	0	0	0	0	0	0
somme	0,3574	0,0875	0,0190	0,0000	0,0000	0	

Tableau 37 : Loi conjointe standard des stations d'Ajaccio et Monaco

La somme des probabilités est égale à 0,46 : elle correspond à la probabilité des phénomènes où les deux stations sont simultanément supérieures à leur espérance respective.

L'analyse de cette loi conjointe standard permet de constater les deux propriétés suivantes :

- De manière générale, toutes les diagonales de la loi sont décroissantes ;
- Les lois marginales de chaque station (qui correspondent à la somme des lignes et des colonnes) est également décroissante.

Ces propriétés confirment les règles de décroissance utilisées pour la génération de tableaux de Young qui sont présentés dans le chapitre V dont les caractéristiques sont :

- décroissance suivant toutes les diagonales ;
- décroissances suivant toutes les lignes et toutes les colonnes.

### Construction de la loi conjointe pour les phénomènes extrêmes

Nous décrivons chaque étape de la construction.

*Etape 1: discrétisation en classes :*

Nous déterminons les 42 combinaisons  $n_k$  en considérant les classes définies précédemment pour les stations d'Ajaccio (7 classes) et Monaco (6 classes).

Le tableau ci-dessous répertorie tous les événements observés pour chaque combinaison de classes.

Classe Ajaccio	1	1	1	1	1	1	2	2	2	2	2	2	3	3	3
Classe Monaco	1	2	3	4	5	6	1	2	3	4	5	6	1	2	3
nk	86	14	0	0	0	0	8	8	3	0	0	0	0	1	1
Classe Ajaccio	3	3	3	4	4	4	4	4	4	5	5	5	5	5	5
Classe Monaco	4	5	6	1	2	3	4	5	6	1	2	3	4	5	6
nk	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
Classe Ajaccio	6	6	6	6	6	6	7	7	7	7	7	7			
Classe Monaco	1	2	3	4	5	6	1	2	3	4	5	6			
nk	0	0	0	0	0	0	0	0	0	0	0	0			

Tableau 38 : Nombre d'événements observés pour chaque combinaison

Etape 2 : Initialisation du taux de risque

- a) Nous devons déterminer le taux de risque  $\lambda_{i,j}$  de chaque combinaison  $n_k$  où  $i$  est l'indice des classes de la station Ajaccio et  $j$  est l'indice des classes de la station Monaco tel que :

$$\lambda_{i,j} \geq \lambda_{i+1,j} \text{ et } \lambda_{i,j} \geq \lambda_{i,j+1}$$

Tous les  $\lambda_{i,j}$  doivent être rangés dans un tableau de Young. Dans ce cas à deux dimensions, il s'agit d'une matrice, de taille 7x6, à l'intérieur de laquelle on a rangé les nombres 1, 2, ..., 42 de telle manière que, sur chaque ligne et chaque colonne, les nombres soient croissants.

Bien entendu, 1 se trouve en haut à gauche et 42 se trouve en bas à droite. Voici un exemple :

1	3	5	19	20	21
2	6	9	22	25	29
4	7	10	24	30	33
8	11	13	27	32	34
12	14	15	31	35	38
16	17	26	36	37	39
18	23	28	40	41	42

Tableau 39 : Exemple d'un des tableaux de Young

- b) On génère  $K = 41$  nombres  $x_k$  selon une loi uniforme entre 0 et 1 et on les réarrange de manière décroissante :  $x_1 \geq x_2 \geq \dots \geq x_{41}$
- c) On considère les 42 "spacings"  $d_k$  :

$$d_1 = x_1, d_2 = x_2 - x_1, \dots, d_{K-1} = x_{K-1} - x_{K-2}, d_K = 1 - x_{K-1}$$

et on réarrange ces « spacings » de manière décroissante.

d) Au final, nous associons chaque  $d_k$  à chaque taux de risque  $\lambda_{i,j}$ , tels qu'ils sont générés dans le tableau de Young.

Classe Ajacc	1	1	1	1	1	1	2	2	2	2	2	2	3	3	3
Classe Monac	1	2	3	4	5	6	1	2	3	4	5	6	1	2	3
$\lambda_k$	0,090	0,063	0,054	0,019	0,018	0,018	0,084	0,054	0,038	0,017	0,016	0,011	0,060	0,050	0,036

Classe Ajacc	3	3	3	4	4	4	4	4	4	5	5	5	5	5	5
Classe Monac	4	5	6	1	2	3	4	5	6	1	2	3	4	5	6
$\lambda_k$	0,016	0,008	0,007	0,039	0,032	0,030	0,014	0,007	0,005	0,031	0,026	0,024	0,007	0,005	0,003

Classe Ajacc	6	6	6	6	6	6	7	7	7	7	7	7
Classe Monac	1	2	3	4	5	6	1	2	3	4	5	6
$\lambda_k$	0,022	0,020	0,015	0,004	0,003	0,002	0,019	0,016	0,012	0,002	0,002	0,001

Tableau 40 : Initialisation des taux de risque de chaque combinaison

L'étape 3 (calcul de la loi de probabilité de l'ensemble  $(p_1, \dots, p_K)$ ) et l'étape 4 (estimation de la loi marginale) sont identiques à celles décrites pour une seule station.

Nous réalisons les 4 étapes en générant un grand nombre (2 000 000) de tableaux de Young. Les valeurs obtenues pour la loi conjointe étendue aux phénomènes extrêmes sont répertoriées dans le tableau ci-dessous :

Ajacc\Monac	]69; 83]	]83; 96]	]96; 110]	]110; 123]	]123; 137]	]137; 150]	somme
]65; 78]	0,1600	0,0289	0,0221	0,0167	0,0095	0,0066	<b>0,2438</b>
]78; 92]	0,0293	0,0207	0,0144	0,0094	0,0080	0,0051	<b>0,0869</b>
]92; 105]	0,0144	0,0101	0,0060	0,0048	0,0046	0,0034	<b>0,0433</b>
]105; 119]	0,0126	0,0081	0,0041	0,0031	0,0026	0,0019	<b>0,0324</b>
]119; 133]	0,0090	0,0071	0,0030	0,0024	0,0018	0,0015	<b>0,0247</b>
]133; 146]	0,0076	0,0060	0,0023	0,0019	0,0010	0,0002	<b>0,0190</b>
]146; 160]	0,0067	0,0036	0,0018	0,0010	0,0005	0,0001	<b>0,0137</b>
somme	<b>0,2397</b>	<b>0,0846</b>	<b>0,0536</b>	<b>0,0392</b>	<b>0,0279</b>	<b>0,0188</b>	

Tableau 41 : Loi conjointe étendue aux phénomènes extrêmes pour les stations d'Ajaccio et de Monaco

Comparons les lois marginales :

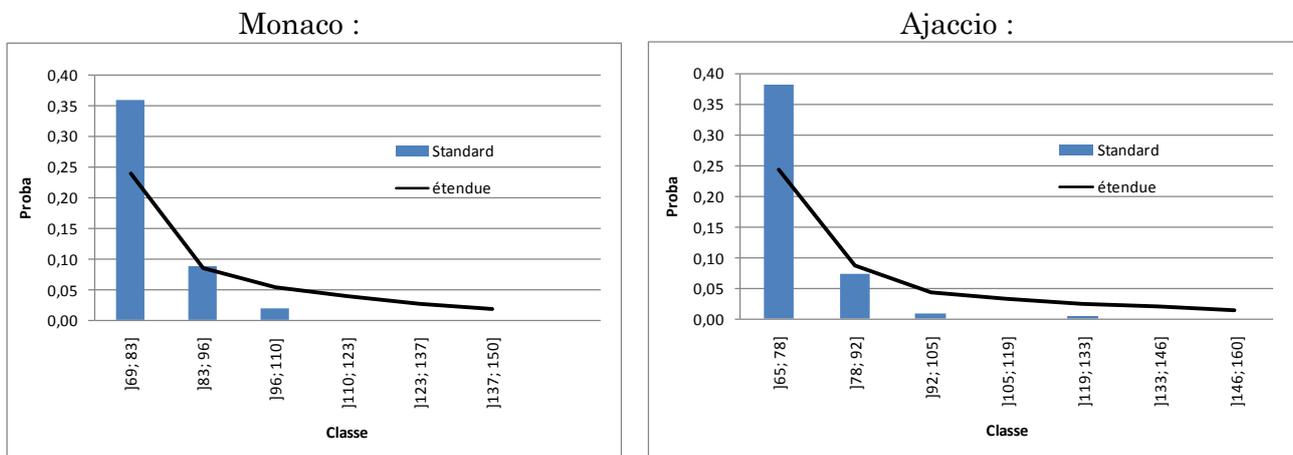


Figure 30 : Lois conjointes standard et étendue aux phénomènes extrêmes des stations de Monaco et d'Ajaccio

Les valeurs obtenues sont critiquables s'il y a une forte dépendance entre stations, car cette dépendance n'a pas été prise en compte à ce stade.

Nous remarquons notamment que la probabilité d'avoir une hauteur de marée très forte dans une station et très faible dans une autre station proche est très certainement beaucoup plus faible que, par exemple, la probabilité d'avoir des valeurs très fortes dans les deux stations simultanément. En effet, les conditions physiques impliquant des hauteurs extrêmes (grandes marées, faible pression atmosphérique, tempêtes...) excluent l'observation de hauteurs très faibles dans une autre station : il existe une dépendance entre les stations d'un même groupe géographique.

Nous allons donc modifier l'approche, pour prendre les dépendances en considération.

Nous introduisons un "masque" que nous déplaçons selon la diagonale. Un masque est un ensemble rectangulaire de cellules, dont les dimensions ne varient pas lorsque le masque est déplacé.

Ensuite, le masque est déplacé selon la diagonale, et la méthode est appliquée séparément pour chaque position du masque. On a donc une application locale de la méthode (à l'intérieur du masque) et non plus une application globale (sur tout le tableau).

Ajacc\Monac	]69; 83]	]83; 96]	]96; 110]	]110; 123]	]123; 137]	]137; 150]	somme
]65; 78]	0,3269962	0,05323194	0	0	0	0	0,3802
]78; 92]	0,03041825	0,03041825	0,01140684	0	0	0	0,0722
]92; 105]	0	0,00380228	0,00380228	0	0	0	0,0076
]105; 119]	0	0	0	0	0	0	0
]119; 133]	0	0	0,00380228	0	0	0	0,0038
]133; 146]	0	0	0	0	0	0	0
]146; 160]	0	0	0	0	0	0	0
somme	0,3574	0,0375	0,0190	0	0	0	

Etape1 : extension de la loi conjointe pour l'ensemble des valeurs contenues ans le masque

Ajacc\Monac	]69; 83]	]83; 96]	]96; 110]	]110; 123]	]123; 137]	]137; 150]	somme
]65; 78]	0,29557863	0,05406109	0,01108146	0	0	0	0,36072118
]78; 92]	0,04097253	0,02890263	0,00731008	0	0	0	0,07718525
]92; 105]	0,00696545	0,00519226	0,00394892	0	0	0	0,01610663
]105; 119]	0,00362904	0,0017916	0,00064235	0	0	0	0,006063
]119; 133]	0	0	0,00380228	0	0	0	0,00380228
]133; 146]	0	0	0	0	0	0	0
]146; 160]	0	0	0	0	0	0	0
somme	0,34714564	0,08994758	0,02678511	0	0	0	

Etape 2 :

Ajacc\Monac	]69; 83]	]83; 96]	]96; 110]	]110; 123]	]123; 137]	]137; 150]	somme
]65; 78]	0,29559025	0,05212485	0,01332776	0	0	0	0,36104286
]78; 92]	0,03910241	0,01656556	0,00778707	0,0034816	0	0	0,06693663
]92; 105]	0,00871322	0,00691099	0,00356927	0,00154621	0	0	0,02073968
]105; 119]	0,0045143	0,00401532	0,00200363	0,00075214	0	0	0,01128538
]119; 133]	0	0,00240024	0,00117601	0,00029753	0	0	0,00387378
]133; 146]	0	0	0	0	0	0	0
]146; 160]	0	0	0	0	0	0	0
somme	0,34792017	0,08201695	0,02786374	0,00607747	0	0	

Etape 3 :

Ajacc\Monac	]69; 83]	]83; 96]	]96; 110]	]110; 123]	]123; 137]	]137; 150]	somme
]65; 78]	0,2964	0,0649	0,0161	0	0	0	0,37750846
]78; 92]	0,0319	0,0149	0,0069	0,0032	0	0	0,05688685
]92; 105]	0,0061	0,0061	0,0022	0,0012	0,0007	0	0,01614959
]105; 119]	0,0036	0,0035	0,0012	0,0006	0,0003	0	0,00926682
]119; 133]	0	0,0021	0,0008	0,0003	0,0001	0	0,00337111
]133; 146]	0	0	0,0005	0,0002	0,0001	0	0,0006955
]146; 160]	0	0	0	0	0	0	0
somme	0,33799823	0,09148613	0,02766975	0,00551327	0,00121094	0	

Etape 4 : obtention de la loi conjointe étendue

Ajacc\Monac	]69; 83]	]83; 96]	]96; 110]	]110; 123]	]123; 137]	]137; 150]	somme
]65; 78]	0,3047	0,0531	0,0112	0	0	0	0,36895565
]78; 92]	0,0358	0,0158	0,0067	0,0031	0	0	0,06140548
]92; 105]	0,0086	0,0067	0,0023	0,0012	0,0007	0	0,01949176
]105; 119]	0,0037	0,0038	0,0013	0,0004	0,0003	0,0002	0,00958307
]119; 133]	0	0,0023	0,0008	0,0003	0,0001	0,0001	0,00353442
]133; 146]	0	0	0,0005	0,0002	0,0001	0,00003	0,00075759
]146; 160]	0	0	0	0,0001	0,00004	0,00001	0,00015037
somme	0,35272624	0,08166756	0,02267183	0,00533358	0,00121281	0,0002663	

Figure 31 : Loi conjointe étendue aux phénomènes extrêmes pour les station d'Ajaccio et de Monaco

La somme des probabilités de tous les tableaux est toujours égale à 0,464.

De cette table, nous pouvons déduire la probabilité de phénomènes conjoints qui n'ont jamais été observés.

Les lois marginales obtenues sont les suivantes :

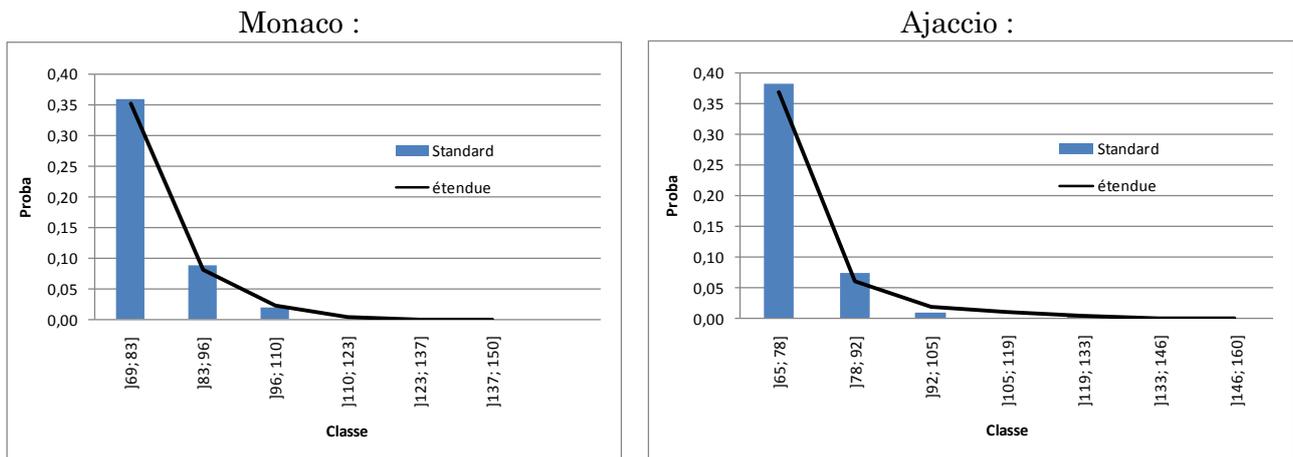


Figure 32 : Lois marginales standards et étendues pour les stations de Monaco et d'Ajaccio

L'utilisation du masque permet d'obtenir des lois marginales étendue plus cohérentes avec les lois de probabilité étendues pour chaque station.

Le choix de la taille du masque à appliquer dépend :

- de la nature des événements à étudier : type d'événements, distance entre les stations, particularité géographique, etc.
- des caractéristiques de la loi standard, notamment du degré de dépendance entre les stations ;
- de la cohérence entre les lois marginales obtenue et les lois de probabilité individuelle de chaque station.

## Références

- [1] Bernard Beuzamy : Robust mathematical methods for extremely rare events, preprint, 2009.
- [2] Bernard Beuzamy : Méthodes probabilistes pour l'étude de phénomènes réels. Ouvrage édité et commercialisé par la *Société de Calcul Mathématique S. A.*, ISBN : 2-9521458-0-6, février 2004.
- [3] Bernard Beuzamy et Olga Zeydina : Méthodes probabilistes pour la reconstruction de données manquantes. Ouvrage édité et commercialisé par la *Société de Calcul Mathématique S. A.*, ISBN : 2-9521458-2-2, ISSN : 1767 – 1175, avril 2007.
- [4] H. Bickert, S. Davin : Méthode probabiliste pour la détermination de la loi conjointe des débits des rivières, Rapport *Société de Calcul Mathématique, S.A.*, 2009.
- [5] Site internet du SHOM (Service Hydrographique et Océanographique de la Marine), [www.sonel.org/ftpWeb/index.php](http://www.sonel.org/ftpWeb/index.php).
- [6] Peter Robinson : Efficient Calculation of Certain Integrals For Modelling Extremely Rare Events, september 2009