



HAL
open science

SYSTÈME DE TRAITEMENT DE DONNÉES MASSIVES APPLIQUÉ À LA SUPERVISION DE MILIEUX CÔTIERS

Dorian D Amico

► **To cite this version:**

Dorian D Amico. SYSTÈME DE TRAITEMENT DE DONNÉES MASSIVES APPLIQUÉ À LA SUPERVISION DE MILIEUX CÔTIERS. Algorithmes et structure de données [cs.DS]. Université de Pau et des Pays de l'Adour, 2021. Français. NNT : 2021PAUU3038 . tel-03596493

HAL Id: tel-03596493

<https://theses.hal.science/tel-03596493v1>

Submitted on 3 Mar 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Systeme de traitement de donnees massives applique à la supervision de milieux côtiers



Dorian d'Amico

Directeurs: **PR. Franck LUTHON** Université de Pau et des Pays de l'Adour
MC. Benoit LARROQUE Université de Pau et des Pays de l'Adour

Examineurs: **PR. Stéphane ABADIE** Université de Pau et des Pays de l'Adour
DR. Xavier BERTIN Université de la Rochelle

Rapporteurs: **DR. Philippe SERGENT** CEREMA Risques Eau Mer
MC HDR. Xavier DESFORGES ENIT - laboratoire LGP

Cette dissertation est soumise pour un
Doctorat en Informatique

15 avril 2021

Remerciements

Je tiens à remercier Monsieur Franck Luthon et Monsieur Benoit Larroque, mes directeurs de thèse, qui m'ont encadré, guidé et écouté tout au long de cette thèse. C'est grâce à eux et à leurs encouragements que cette thèse a pu aboutir.

J'adresse tous mes remerciements à Monsieur Philippe Sergent ainsi qu'à Monsieur Xavier Desforges de l'honneur qu'ils m'ont fait en acceptant d'être rapporteurs de cette thèse et d'avoir pris le temps de lire et de commenter mes travaux.

J'exprime ma gratitude à Monsieur Xavier Bertin et à Monsieur Stéphane Abadie qui ont bien voulu être examinateurs lors de ma soutenance de thèse.

Je veux aussi remercier les membres du LIUPPA pour toutes les discussions et les conseils qui m'ont permis d'avancer dans mes réflexions.

Je souhaite également remercier Monsieur Stéphane Abadie ainsi que tous les membres de l'équipe SIAME avec qui j'ai eu l'opportunité de travailler et d'échanger. Leurs connaissances m'ont permis d'établir un cadre dans mes recherches et d'avancer dans la bonne direction.

Aussi, je remercie ma famille qui m'a toujours apporté son soutien et grâce à qui j'ai pu arriver jusqu'ici.

Enfin, je remercie ma conjointe sur qui j'ai pu me reposer afin d'avancer sereinement tout au long de mes travaux et qui m'a aidé durant ces trois années en me relisant et en m'aidant à préparer mes présentations orales.

Résumé

La puissance des vagues a pour effet néfaste d'endommager les côtes et les ouvrages qui les protègent. Les réparations qui s'ensuivent sont très coûteuses. C'est pour cette raison que le déferlement des vagues est étudié depuis un siècle, dans le but de mieux comprendre ce phénomène complexe, d'améliorer les structures côtières et de mieux superviser ces ouvrages. Cependant, l'environnement hostile de l'océan a longtemps empêché de réaliser des mesures sur le terrain sur de longues périodes et notamment lors de tempêtes. De ce fait, beaucoup d'études ont été réalisées en laboratoire, permettant la compréhension de beaucoup de mécanismes. Toutefois, ces études ne représentent pas toujours la réalité. En effet, en laboratoire l'échelle est différente, les fluides utilisés sont différents, il n'y a pas de perturbations liées aux paramètres environnementaux (vent, courant, etc.) et le phénomène est en 2D alors qu'en cas réel il est en 3D. Il est donc nécessaire, afin d'avoir une meilleure compréhension des impacts de vagues sur les structures côtières, d'étudier ce phénomène sur le terrain à échelle réelle, là où les conditions ne sont pas configurables mais bien dépendantes de l'environnement. Le manque de données limite grandement les études possibles. Néanmoins, grâce à la technologie moderne, nous avons mis en place une station de mesure *in-situ*, sur la digue de l'Artha à Saint-Jean-de-Luz, en France. Nos recherches se basent sur le grand nombre de données enregistrées par cette station. Le premier résultat a été de mettre en place un algorithme permettant de détecter les impacts présents dans les signaux enregistrés. Ensuite, afin de caractériser ces impacts, différents critères ont été déterminés et sont calculés automatiquement. Ces critères ont été utilisés afin de prédire les pressions d'impact des vagues sur la digue de l'Artha grâce aux paramètres environnementaux locaux, en utilisant l'apprentissage automatique et plus spécifiquement les réseaux de neurones. Les facteurs environnementaux ne semblant pas suffisants, nous avons étudié les types d'impacts, qui jouent un rôle majeur dans la pression générée par une vague sur un mur. A cause des différences d'échelles de pressions et de temps, il n'a pas été possible de réaliser une classification reposant sur les caractéristiques des types d'impact établis en laboratoire. De ce fait, nous avons utilisé quatre algorithmes d'apprentissage non-supervisé, que nous avons comparé afin de réaliser un *clustering* des impacts de vagues. Deux algorithmes ont donné des résultats proches, un troisième a donné un résultat différent, et le dernier n'a pas fonctionné.

Mots clés : Pression d'impact, Impact de vague, Type d'impact, Digue, Littoral, Détection d'impact de vague, Station de mesure, Mesure en continu, Mesure in-situ, Machine Learning, Apprentissage automatique, Apprentissage supervisé, Apprentissage non-supervisé.

Table des matières

Table des figures	xv
Liste des tableaux	xxiii
Introduction	1
1 Contexte	3
1.1 Les vagues	4
1.1.1 Formation des vagues	4
1.1.2 Facteurs de formation des vagues	5
1.1.3 Caractéristiques d'une vague	6
1.1.4 Dispersion des vagues	7
1.1.5 Déferlement des vagues	8
1.2 Les structures côtières	10
1.2.1 Les différents types de structures	10
1.2.1.1 Epis	10
1.2.1.2 Diguees	11
1.2.1.3 Brise-lames	13
1.2.1.4 Jetées	15
1.2.2 Effet de la houle sur les structures	16

1.3	Objectifs	17
2	Etat de l'art	19
2.1	Pression d'impact des vagues	20
2.1.1	Etudes in-situ	21
2.1.1.1	De Rouville	21
2.1.1.2	Bird et Bullock	24
2.1.2	Etudes en laboratoire	24
2.1.2.1	Hofland	24
2.1.2.2	Kaminski	27
2.1.3	Bilan	29
2.2	Apprentissage	30
2.2.1	Exploration de données	30
2.2.1.1	Qu'est-ce que l'exploration de données ?	30
2.2.1.2	Ensemble de données	30
2.2.1.3	Knowledge Discovery in Databases	31
2.2.2	Standardisation des entrées	32
2.2.3	Apprentissage supervisé	33
2.2.3.1	Réseaux de neurones	33
2.2.3.2	Machine à vecteurs de support	39
2.2.4	Apprentissage non supervisé	41
2.2.4.1	K-Means	41
2.2.4.2	Self-Organizing Maps	43
2.2.4.3	Density-Based Spatial Clustering of Applications with Noise	44
2.2.4.4	Hierarchical Clustering	46
2.2.5	Classification et clustering	49

2.2.6	K-fold cross validation	50
2.2.7	Bilan	51
3	Expérimentation	53
3.1	Besoins	54
3.2	Station de l'Artha	55
3.2.1	Histoire de la baie de Saint-Jean-de-Luz	55
3.2.2	La station	59
3.2.2.1	Les capteurs	59
3.2.2.2	Station de mesure	62
3.2.2.3	Stockage	63
3.2.2.4	Système de transmission	63
3.2.2.5	Alimentation	64
3.2.2.6	Algorithme d'acquisition	65
3.2.2.7	Système d'interactions embarqué	67
3.2.2.8	Station dans son ensemble	67
3.2.2.9	Mise en place de la station sur la digue de l'Artha	70
3.2.2.10	Données de pression de la station de l'Artha	71
3.2.3	Données provenant d'autres stations de mesure	72
3.2.3.1	Données de la station météo de Socoa	73
3.2.3.2	Données du marégraphe de Socoa	75
3.2.3.3	Données du houlographe	75
3.3	Bilan	77
4	Résultats	79
4.1	Prétraitement des signaux de la station de l'Artha	80
4.1.1	Signaux bruts	80

4.1.2	Conversion des données	81
4.1.3	Filtrage non-linéaire	81
4.1.4	Filtrage linéaire	83
4.1.4.1	Filtre de Butterworth	84
4.1.4.2	Filtre de Tchebychev	84
4.1.5	Temps de calcul	86
4.2	Détection automatique des impacts	87
4.2.1	Objectif	87
4.2.2	Algorithmes	88
4.2.2.1	Approche 1 : Algorithme sans paramètres empiriques	88
4.2.2.2	Approche 2 : Algorithme basé sur la dérivée	90
4.2.2.3	Approche 3 : Algorithme basé sur la suppression de la composante continue	91
4.2.2.4	Conclusion	93
4.3	Calcul de critères d'impact	95
4.3.1	Critères temporels	95
4.3.1.1	Temps de montée	95
4.3.1.2	Temps de descente	95
4.3.1.3	Durée	96
4.3.2	Critères de pression	97
4.3.2.1	Maximum	97
4.3.3	Intégrale	97
4.3.4	Impulsion de pression et durée d'impulsion	98
4.3.5	Force linéique maximale	100
4.3.6	Impacts associés	102
4.3.7	Bases de données	104

4.3.7.1	Base de données par impact	104
4.3.7.2	Base de données par enregistrement	105
4.4	Calcul de covariance entre les capteurs	107
4.5	Visualisation animée de la pression d'impact sur les capteurs	110
4.5.1	Evolution linéaire de la pression d'impact	110
4.5.2	Evolution 2D de la pression d'impact	112
4.6	Retour sur la station de l'Artha	114
4.7	Prévision de la pression d'impact des vagues	116
4.7.1	Etape 1 : Objectif	116
4.7.2	Etape 2 : Les données	116
4.7.3	Etape 3 : Prétraitement des données	117
4.7.4	Etape 4 : Format des données	118
4.7.5	Choix de l'approche	119
4.7.6	Choix de l'algorithme	120
4.7.7	Etape 7 : Application de l'algorithme	121
4.7.8	Etape 8 : Interprétation des résultats	122
4.7.9	Conclusion	123
4.8	Clustering des types d'impacts des vagues	125
4.8.1	Etape 1 : Objectif	125
4.8.2	Etape 2 : Les données	127
4.8.3	Etape 3 : Prétraitement des données	127
4.8.4	Etape 4 : Format des données	128
4.8.5	Etape 5 : Choix de l'approche	128
4.8.6	Etape 6 : Choix de l'algorithme	129
4.8.7	Etape 7 : Application de l'algorithme	129
4.8.7.1	Paramétrage Hierarchical clustering	129

4.8.7.2	Paramétrage SOM	130
4.8.7.3	Paramétrage DBSCAN	130
4.8.8	Etape 8 : Interprétation des résultats	131
4.8.8.1	Résultats Hierarchical clustering	131
4.8.8.2	Résultats SOM	139
4.8.8.3	Résultats DBSCAN	149
4.8.9	Comparaison des clusters	150
4.8.10	K-Means	153
4.8.11	Conclusion	161
Conclusion et perspectives		163
Bibliographie		165
Publications et conférences		171
Annexe A Extraits des tableaux de données		173
A.1	Tableaux des données environnementales	173
A.1.1	Extrait des données de la station météo de Socoa	173
A.1.2	Extrait des données du marégraphe de Socoa	176
A.1.3	Extrait des données du houlographe	177
A.2	Bases de données réalisées	179
A.2.1	Base de données d'impacts	179
A.2.2	Base de données par enregistrements	182
Annexe B Tables de données liées au réseaux de neurones de la section 4.7		183
B.1	Format des données utilisé pour le réseau de neurones	183
B.2	Tableau des données de validation du réseau de neurones	186

Table des matières	xiii
--------------------	-------------

Annexe C Photos de la mise en place de la station de la digue de l'Artha	209
---	------------

Annexe D Fiche technique du capteur de pression	219
--	------------

Table des figures

1.1	Schéma qui représente la formation de la houle [2].	5
1.2	Caractéristiques d'une vague.	6
1.3	Mouvement orbital des particules en eau profonde [4].	7
1.4	Mouvement orbital des particules dans l'eau de profondeur intermédiaire [4].	8
1.5	Mouvement orbital des particules en eau peu profonde [4].	8
1.6	Types de déferlement des vagues. De gauche à droite : spilling, plunging and surging [11].	9
1.7	Exemple d'implantation d'épis où un mécanisme d'érosion/accrétion autour de l'ouvrage est observable [13].	11
1.8	Exemple d'une digue sur une plage [17].	13
1.9	Exemple d'un brise-lame en reconstruction [19].	15
1.10	Exemple d'une longue jetée [20].	16
1.11	Brise-lames détruit à Takenoura, Japon, lors du tsunami en 2011 [21]. . . .	17
2.1	Types d'impact des vagues en fonction du déferlement [32].	26
2.2	Exemple de génération de vague dans le canal de Delta [32].	27
2.3	Signal d'impact pour chaque type d'impact à différentes hauteur d'une paroi [37].	28
2.4	Extrait d'un ensemble de données.	31
2.5	Standardisation des features.	33

2.6	Schéma simplifié de deux neurones biologiques [45].	34
2.7	Architecture d'un réseau de neurones [48].	36
2.8	Structure d'un neurone artificiel.	37
2.9	Fonctions d'activation. A) Seuil ; B) Linéaire ; C) Sigmoidé ; D) Redresseur ; E) Tanh.	38
2.10	Hyperplans dans l'espace 2D et 3D [52].	40
2.11	Hyperplan optimal [52].	40
2.12	Illustration de clustering avec K-Means [55].	42
2.13	Illustration de la formation d'une SOM [59].	44
2.14	Schéma des différents points de l'algorithme DBSCAN avec $minPts = 4$ [63].	45
2.15	Schéma du processus de clustering hiérarchique.	47
2.16	Exemple de Dendrogramme.	48
2.17	Exemple de dendrogramme avec ligne de séparation des clusters.	49
2.18	Schéma de <i>K-Fold Cross Validation</i> avec $k = 10$	51
3.1	Photo de la baie de Saint-Jean-de-Luz où les trois digues sont présentes. La première digue est la digue de Sainte-Barbe, celle au milieu de l'eau est la digue de l'Artha, et enfin celle du fond est la digue de Socoa.	56
3.2	Photos de 1951 montrant les dégâts de la tempête à Saint-Jean-de-Luz [77].	57
3.3	Photo de 1951 montrant un bloc de 40 tonnes sur la digue [77].	57
3.4	Photo du 28 février 2017 montrant une vague déposant un bloc de 50 tonnes sur la digue de l'Artha [78].	58
3.5	Photo de mars 2017 montrant le bloc de 50 tonnes sur la digue de l'Artha [78].	58
3.6	Schéma d'un capteur piézorésistif [79].	59
3.7	Caractéristiques principales du capteur utilisé.	60
3.8	Schéma du positionnement des capteurs de pression.	61
3.9	Photo rapprochée du mur avec les capteurs. Le capteur 0 est entouré en rouge.	62

3.10	Photo panoramique du musoir ouest avec les capteurs.	62
3.11	Photo de la digue de l'Artha où se situe la station et le panneau photovoltaïque. <i>N.B.</i> : Les capteurs sont sur l'autre pan de mur du musoir peint en blanc. . .	64
3.12	Algorigramme de l'algorithme d'acquisition des mesures de pression. . . .	66
3.13	Schéma synthétisant le logiciel embarqué permettant de communiquer à distance avec la station. CMU : <i>Central Measurement Unit</i> ; <i>Fs</i> : fréquence d'échantillonnage; <i>Tacq</i> : durée d'une acquisition.	67
3.14	Schéma de la configuration de l'instrumentation pour l'acquisition et la transmission de données autonomes (CMU : <i>Central Measurement Unit</i> ; VPN : <i>Virtual Private Network</i> ; PV : panneau photovoltaïque).	69
3.15	Photo du panneau solaire et de l'installation électrique intégrée dans la digue.	70
3.16	Digue de l'Artha avec l'emplacement des capteurs indiqué par un point rouge. Les étoiles rouges correspondent à (A) Houlographe 06403; (B) Marégraphe de Socoa; (C) Station météo de Socoa [83].	73
4.1	Signal brut en μA d'un capteur.	80
4.2	Signal converti en bar d'un capteur.	81
4.3	Signal avant application du filtre médian contenant une valeur aberrante. . .	82
4.4	Signal après application du filtre médian.	83
4.5	Réponse en fréquence du filtre de Butterworth.	84
4.6	Réponse en fréquence du filtre de Tchebyshev.	85
4.7	Signal filtré d'un capteur.	86
4.8	Signal d'impact contenant plusieurs pics.	88
4.9	Signal contenant un décalage de la composante continue.	89
4.10	Dérivée d'une partie d'un signal avec seuils de détections.	90
4.11	Pression atmosphérique entre chaque impact.	92
4.12	Signal avant et après application d'un filtre passe-haut.	92
4.13	Temps de montée de 10% à 90% du maximum.	95

4.14	Temps de descente de 90% à 10% du maximum.	96
4.15	Durée de 10% du maximum avant le pic à 10% du maximum après le pic.	96
4.16	Maximum absolu et maximum relatif d'un impact.	97
4.17	Intégrale de 10% du maximum avant le pic à 10% du maximum après le pic.	98
4.18	Impulsion de pression et durée d'impulsion.	99
4.19	Impact sans impact percussif.	99
4.20	Force linéique et pression d'impact sur les 11 capteurs de l'axe vertical	101
4.21	Inclusion temporelle d'un impact dans un autre.	103
4.22	Intersection temporelle entre deux impacts.	103
4.23	Covariance sur deux impacts.	107
4.24	Covariance sur une partie de signal comprenant plusieurs impacts.	108
4.25	Spectre des signaux et spectre de la covariance.	109
4.26	Impact d'une vague sur les 16 capteurs de la digue de l'Artha	110
4.27	Capture de la vidéo à 0.86 seconde : graphique permettant la visualisation linéaire verticale de la pression sur les capteurs	111
4.28	Capture de la vidéo à 0.52 seconde : graphique permettant la visualisation 2D de la pression sur les capteurs. La valeur à côté de chaque capteur correspond à la pression sur le capteur en bar. Le code couleur de vert à rouge permet d'avoir une indication de la force de la pression sur chaque capteur.	113
4.29	Histogramme montrant le nombre d'impacts par capteur entre juin 2018 et juin 2019.	114
4.30	Histogramme montrant le nombre d'impacts supérieurs à 0.3 bar en relatif par capteur entre juin 2018 et juin 2019.	115
4.31	Photo du profil de la digue de l'Artha.	115
4.32	Rose des vents à 16 directions.	117
4.33	Graphique représentant l'incidence de chaque variable environnementale sur la pression d'impact des vagues (en pourcentage).	120

4.34	Histogramme représentant les erreurs du modèle par intervalles. Valeur des erreurs en abscisses et nombre d'exemples en ordonnées.	123
4.35	Enregistrement de 10 minutes contenant la pression d'impact la plus puissante.	126
4.36	Pression d'impact la plus puissante enregistrée.	126
4.37	Dendrogramme représentant le clustering des impacts.	131
4.38	Tableau représentant les intervalles pour chaque <i>feature</i> de chaque cluster du dendrogramme.	132
4.39	Graphique représentant les intervalles de chaque critère par cluster obtenu avec Hierarchical clustering.	133
4.40	Echantillons de signaux d'impact du cluster 0 du dendrogramme.	134
4.41	Echantillons de signaux d'impact du cluster 1 du dendrogramme.	134
4.42	Echantillons de signaux d'impact du cluster 2 du dendrogramme.	135
4.43	Echantillons de signaux d'impact du cluster 3 du dendrogramme.	135
4.44	Echantillons de signaux d'impact du cluster 4 du dendrogramme.	136
4.45	Signal d'impact type moyenné pour le cluster 0 obtenu avec l'algorithme Hierarchical clustering.	137
4.46	Signal d'impact type moyenné pour le cluster 1 obtenu avec l'algorithme Hierarchical clustering.	137
4.47	Signal d'impact type moyenné pour le cluster 2 obtenu avec l'algorithme Hierarchical clustering.	138
4.48	Signal d'impact type moyenné pour le cluster 3 obtenu avec l'algorithme Hierarchical clustering.	138
4.49	Signal d'impact type moyenné pour le cluster 4 obtenu avec l'algorithme Hierarchical clustering.	139
4.50	Carte de dimension 16 par 16 obtenue avec l'algorithme SOM.	140
4.51	Carte de dimension 4 par 4 obtenue avec l'algorithme SOM.	141
4.52	Clusters déterminés par l'analyse de la carte de dimension 4 par 4 obtenue avec l'algorithme SOM.	142

4.53	Tableau représentant les intervalles pour chaque <i>feature</i> de chaque cluster du SOM.	143
4.54	Graphique représentant les intervalles de chaque critère par cluster obtenu avec SOM.	144
4.55	Echantillons de signaux d'impact du cluster 0 du SOM.	145
4.56	Echantillons de signaux d'impact du cluster 1 du SOM.	145
4.57	Echantillons de signaux d'impact du cluster 2 du SOM.	146
4.58	Echantillons de signaux d'impact du cluster 3 du SOM.	146
4.59	Echantillons de signaux d'impact du cluster 4 du SOM.	147
4.60	Signal d'impact type moyenné pour le cluster 0 obtenu avec l'algorithme SOM.	147
4.61	Signal d'impact type moyenné pour le cluster 1 obtenu avec l'algorithme SOM.	148
4.62	Signal d'impact type moyenné pour le cluster 2 obtenu avec l'algorithme SOM.	148
4.63	Signal d'impact type moyenné pour le cluster 3 obtenu avec l'algorithme SOM.	149
4.64	Signal d'impact type moyenné pour le cluster 4 obtenu avec l'algorithme SOM.	149
4.65	Clustering obtenu avec l'algorithme DBSCAN.	150
4.66	Comparaison des impacts en commun entre les clusters des deux algorithmes.	151
4.67	Tableau représentant les intervalles pour chaque <i>feature</i> de chaque cluster de K-Means.	154
4.68	Graphique représentant les intervalles de chaque critère par cluster obtenu avec K-Means.	154
4.69	Echantillons de signaux d'impact du cluster 0 de K-Means.	155
4.70	Echantillons de signaux d'impact du cluster 1 de K-Means.	155
4.71	Echantillons de signaux d'impact du cluster 2 de K-Means.	156
4.72	Echantillons de signaux d'impact du cluster 3 de K-Means.	156
4.73	Echantillons de signaux d'impact du cluster 4 de K-Means.	157
4.74	Signal d'impact type moyenné pour le cluster 0 obtenu avec l'algorithme K-Means.	158

4.75	Signal d'impact type moyenné pour le cluster 1 obtenu avec l'algorithme K-Means.	158
4.76	Signal d'impact type moyenné pour le cluster 2 obtenu avec l'algorithme K-Means.	159
4.77	Signal d'impact type moyenné pour le cluster 3 obtenu avec l'algorithme K-Means.	159
4.78	Signal d'impact type moyenné pour le cluster 4 obtenu avec l'algorithme K-Means.	160
4.79	Comparaison des impacts en communs entre <i>Hierarchical Clustering</i> et K-Means.	161
A.1	Début d'un extrait de données de la station météo de Socoa.	174
A.2	Suite d'un extrait de données de la station météo de Socoa.	175
A.3	Extrait de données provenant du marégraphe de Socoa.	176
A.4	Début d'un extrait de données du houlographe de Saint-Jean-de-Luz.	177
A.5	Suite d'un extrait de données du houlographe de Saint-Jean-de-Luz.	178
A.6	Début de l'extrait de la base de données des impacts	180
A.7	Suite de l'extrait de la base de données des impacts	181
A.8	Extrait de la base de données par enregistrement et par capteur	182
B.1	Début d'un extrait de données formatées.	184
B.2	Suite d'un extrait de données formatées.	185
C.1	Photo d'un agent du CD64 en train de réaliser la saignée sur la digue à l'aide d'un marteau piqueur.	210
C.2	Photo de deux agents du CD64 en train de remettre en état l'endroit où sera placé la station.	211
C.3	Photo de deux agents du CD64 finalisant la mise en place des caissons sur lesquels les plaques des capteurs vont être fixées.	212
C.4	Photo des deux plaques sur lesquelles les capteurs sont fixés.	213

C.5	Photo de la mise en place de la gaine et tirage des câbles des capteurs. . . .	213
C.6	Coulage du béton sur la gaine électrique où passent les câbles des capteurs.	214
C.7	Mise en place de la plaque horizontale de capteurs, vue de haut.	215
C.8	Mise en place de la plaque verticale de capteurs, vue de haut.	216
C.9	Mise en place de la plaque verticale de capteurs.	217

Liste des tableaux

B.1	Tableau présentant les résultats du réseau de neurones sur le jeu de test.	. . . 207
-----	--	-----------

Glossaire

cluster groupe de données ayant des similarités. 49

DBSCAN Density-Based Spatial Clustering of Applications with Noise, méthode de clustering. 44

DD direction moyenne du vent. 74

DropTimeMax temps de descente du signal d'impact entre le maximum et la fin en seconde. 128

Duration durée de l'impact en seconde. 128

fetch distance ininterrompue de l'eau libre sur laquelle le vent souffle sans changement significatif de direction. 5

Ff vitesse moyenne du vent. 74

h110d hauteur moyenne du dixième supérieur en hauteur des vagues. 75

HC Hierarchical Clustering, méthode de clustering. 150

ImpulseDuration durée de l'impulsion en seconde. 128

Integrale intégrale du signal d'impact. 128

KDD Knowledge Discovery in Databases. 31

K-Means méthode de clustering. 41

ML Machine Learning. 120

P1/10 moyenne des 10% plus grands pics de pression de l'enregistrement en bar. 106

PressureImpulse impulse de pression. 128

RelativeMaxValue valeur relative du maximum de l'impact en bar. 128

RiseTimeMax temps de montée du signal d'impact entre le début et le maximum en seconde. 128

SOM Self-Organizing Maps, méthode de clustering. 43

th110d moyenne des périodes du dixième supérieur en hauteur des vagues. 76

thetap direction de provenance au pic de la vague. 76

Introduction

Le changement climatique que nous vivons a de nombreuses conséquences sur notre planète. En effet, si nous nous concentrons seulement sur les océans, le réchauffement climatique engendre au moins le réchauffement des eaux, la montée des eaux et de fortes tempêtes. Ces changements impactent la vie, que ce soit dans les océans ou sur terre.

Dans les océans, ces modifications provoquent la disparition progressive de certains écosystèmes, comme les récifs de la Grande Barrière (*Great Barrier Reef*) en Australie par exemple. Ce récif est le plus grand récif corallien du monde et il abrite beaucoup d'espèces en tout genre. Cependant, ces coraux souffrent de maladies dont une des principales causes est le réchauffement climatique actuel. La disparition de ces coraux provoquera probablement la disparition de la biodiversité qui l'accompagne.

En ce qui concerne les effets du changement climatique sur terre, la montée des eaux et les fortes tempêtes impactent grandement les littoraux. En effet, les phénomènes extrêmes sont de plus en plus fréquents et leurs effets ont souvent de grandes conséquences. D'un point de vue humain, ils provoquent de nombreux décès et obligent des grands mouvements de population. D'un point de vue économique, une partie du littoral touché peut être dévastée, engendrant de grosses pertes et nécessitant une reconstruction massive.

Il serait donc intéressant d'être capable de prévoir de tels événements extrêmes et d'adapter les structures côtières afin de pouvoir y faire face. Pour cela, il faut être capable de suivre et surveiller l'action des vagues sur les côtes, et plus spécifiquement sur les structures côtières qui protègent le littoral. Dans le travail présenté dans ce manuscrit, nous nous concentrons sur ce dernier point.

Notre objectif est donc d'étudier la pression d'impact des vagues sur les structures côtières afin d'essayer de mieux comprendre ce phénomène complexe et de mieux superviser ces structures. Pour cela, le document est divisé en quatre chapitres.

Dans le chapitre 1 nous verrons le contexte de ce travail dans sa globalité. Nous allons parler des vagues de manière générale. Ensuite, nous présenterons les différentes structures côtières et leur utilité. Nous finirons ce premier chapitre en fixant les objectifs du travail présenté ici.

Le chapitre 2 présentera dans un premier temps un état de l'art sur l'étude de la puissance d'impact des vagues. Pour cela nous présenterons différentes études qui ont été réalisées et les résultats principaux qui sont intéressants pour notre travail. Dans un second temps, nous présenterons un état de l'art sur l'apprentissage, en présentant les méthodes principales d'apprentissage supervisé et d'apprentissage non-supervisé.

Le chapitre 3 présentera les prérequis nécessaires pour réaliser le travail que nous souhaitons faire. Ensuite, nous présenterons en détail la station que nous avons mise en place afin de collecter des données de puissances d'impact de vagues sur de longues périodes.

Le chapitre 4 quant à lui présentera les pistes explorées et les résultats obtenus.

Enfin, une conclusion permettra de faire un bilan du travail réalisé et des résultats obtenus.

Chapitre 1

Contexte

Ce sujet de thèse fait appel à des connaissances en informatique et en ingénierie côtière. Il est donc nécessaire de commencer par contextualiser le projet afin de pouvoir comprendre les objectifs de ce travail.

Pour cela, nous allons voir dans ce chapitre quelques notions générales permettant de mieux appréhender le reste des travaux. Nous commencerons par aborder le sujet des vagues, la manière dont elles se forment, la façon dont elles se propagent et quelles sont leurs caractéristiques.

Ensuite, nous présenterons les différentes structures côtières qui sont fréquemment utilisées, après quoi nous parlerons des effets de la houle sur ces structures.

Enfin, nous présenterons les objectifs de cette thèse.

1.1 Les vagues

En anglais, le mot vague se dit *wave*. Dans cette langue, ce mot veut également dire onde, ce qui est le mot juste qui caractérise une vague. En effet, une vague n'est rien d'autre qu'une onde spécifique, qui se propage dans l'eau.

Il existe plusieurs types d'ondes de vagues. Ces différents types d'ondes sont principalement caractérisés par deux facteurs : la période et la source génératrice des ondes. Sans entrer dans les détails, ces différentes ondes sont les suivantes [1], [2] :

- Ondes capillaires,
- Ondes ultra-gravitationnelles,
- Ondes de gravité ordinaires,
- Ondes infra-gravitationnelles,
- Vagues de longue période,
- Marées ordinaires,
- Raz-de-marée.

Dans ce travail nous allons nous intéresser qu'aux ondes de gravité ordinaires, qui sont les vagues classiques que nous avons l'habitude d'observer au bord de la plage.

1.1.1 Formation des vagues

L'origine des vagues est une source d'énergie qui est transmise à l'eau et qui engendre une onde. Cette source d'énergie provient le plus souvent du vent. En effet, les vagues de vent sont créées par la friction entre le vent et la surface de l'eau. Quand le vent souffle à la surface de l'océan ou d'un lac, la perturbation continue crée une crête de vague. Ces types de vagues se trouvent partout dans les océans et le long des côtes.

Cependant il existe également d'autres sources d'énergie, plus rares, qui génèrent d'autres types de vagues. Ces vagues plus rares sont causées par des intempéries, comme des ouragans. Les vents forts et la forte pression de ce type de tempête causent des vagues de tempête, une série de longues vagues qui sont créées loin de la rive dans les eaux plus profondes. D'autres vagues dévastatrices peuvent être provoquées par des perturbations sous-marines qui déplacent rapidement de grandes quantités d'eau, telles que des tremblements de terre, des glissements de terrain ou des éruptions volcaniques. Cela crée de très longues vagues, qui sont appelées des tsunamis [3].

Une fois que le vent a cessé de souffler, les vagues générées par le vent sont appelées "houle". Plus généralement, ce que nous appelons houle sont des vagues générées par le vent, qui ne sont plus significativement affectées par le vent local à ce moment-là. Ces vagues ont

été générées ailleurs [4]. Une vague est donc caractérisée comme une onde qui se propage à la surface de l'eau entre deux fluides, l'eau et l'air.

Les vagues ont une part de hasard. En effet, des vagues subséquentes diffèrent en hauteur, en durée et en forme avec une prévisibilité limitée. Elles peuvent être décrites comme un processus stochastique, en combinaison avec la physique qui régit leur génération, leur croissance et leur propagation.

1.1.2 Facteurs de formation des vagues

Cinq facteurs influencent la formation des vagues générées par le vent [5] :

- La vitesse ou la force du vent par rapport à la vitesse de la vague (le vent doit se déplacer plus vite que la crête de la vague pour lui transférer de l'énergie),
- Le fetch, ce qui correspond à la distance ininterrompue de l'eau libre sur laquelle le vent souffle sans changement significatif de direction,
- La largeur de la zone affectée par le fetch,
- La durée pendant laquelle le vent a soufflé sur l'eau,
- La profondeur de l'eau.

Tous ces facteurs fonctionnent ensemble afin de déterminer les caractéristiques des vagues. La figure 1.1 schématise la création de la houle par le vent en mer.

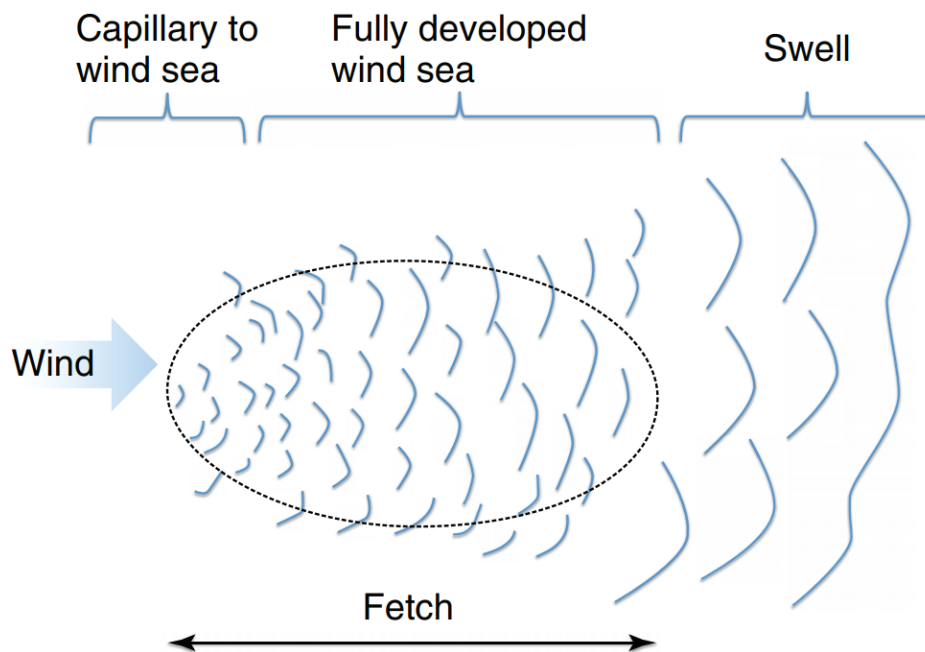


FIGURE 1.1 Schéma qui représente la formation de la houle [2].

La hauteur d'une vague est déterminée par la vitesse du vent, la durée du vent et le fetch. Si la vitesse du vent est lente, alors celui-ci générera de petites vagues. Si le vent est puissant mais pendant une courte période, alors il n'y aura pas de grosses vagues. Si le vent est fort pendant une longue durée mais sur un fetch de petite taille, il n'y aura pas non plus de grosses vagues. D'après Duxbury *et al.* [6], pour qu'il y ait de grosses vagues, les trois facteurs doivent être combinés.

1.1.3 Caractéristiques d'une vague

Les principales caractéristiques des vagues sont les suivantes :

- La hauteur H (distance verticale du creux à la crête),
- L'amplitude A (distance entre le niveau de l'eau au repos et la crête),
- La longueur d'onde L (distance crête à crête dans la direction de la propagation),
- La période d'onde P (intervalle de temps entre l'arrivée des crêtes consécutives à un point stationnaire),
- La direction de propagation S ,
- La célérité C ($\frac{L}{P}$),
- La profondeur D .

La figure 1.2 ci-dessous illustre ces caractéristiques.

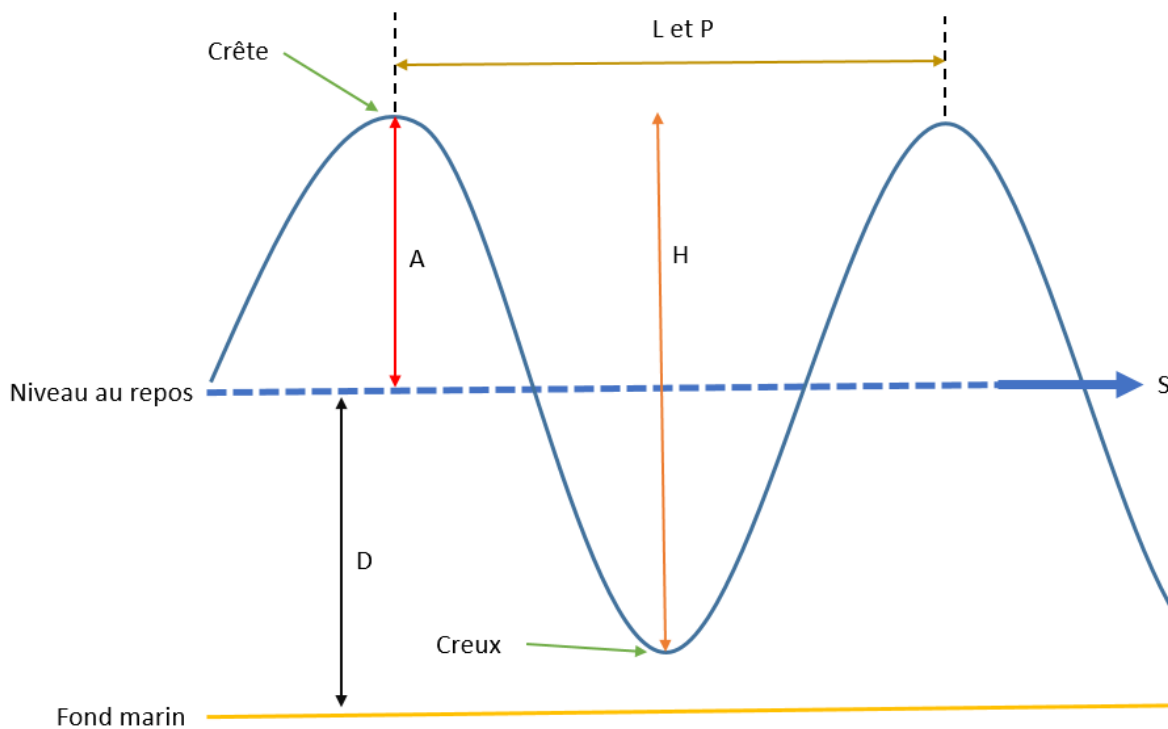


FIGURE 1.2 Caractéristiques d'une vague.

La célérité d'une vague peut être calculée grâce à la formule suivante : $\frac{L}{P}$. Il est intéressant de connaître la célérité d'une vague. Cependant, dans certains cas il n'est pas possible de la calculer à cause d'un manque d'information, comme la longueur d'onde. Il nous est donc impossible de calculer la célérité. Mais grâce à l'équation de dispersion de vagues qu'a défini Guo *et al.* [7], il est possible d'estimer la longueur d'onde et donc de calculer la célérité. Cela nécessite de connaître la bathymétrie de la zone de propagation des vagues étudiées, car la dispersion dépend de la profondeur.

1.1.4 Dispersion des vagues

Lorsqu'une vague se propage, elle ne transporte pas d'eau sur son passage. En effet, lorsque la vague passe, elle fait bouger les particules d'eau dans un mouvement circulaire. Ce mouvement est différent en fonction de la profondeur. En eau profonde, l'eau décrit un mouvement parfaitement circulaire, qui est de plus en plus petit avec la profondeur. La figure 1.3 décrit ce phénomène.

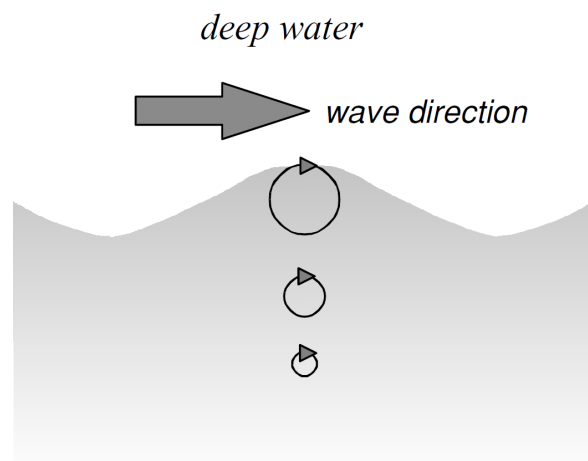


FIGURE 1.3 Mouvement orbital des particules en eau profonde [4].

Cependant, en eau moins profonde, plus on s'approche du fond, plus l'eau circule dans un mouvement elliptique. Les figures 1.4 et 1.5 montrent cela, respectivement en eau moyennement profonde puis en eau peu profonde.

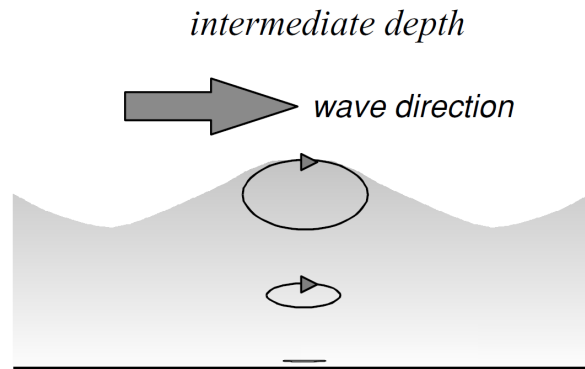


FIGURE 1.4 Mouvement orbital des particules dans l'eau de profondeur intermédiaire [4].

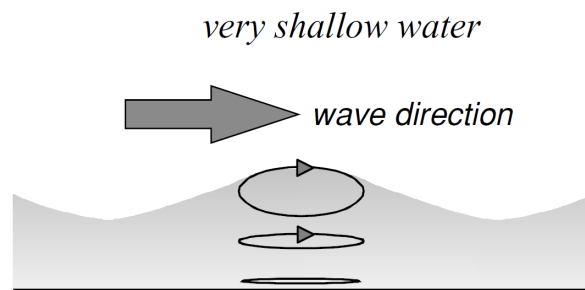


FIGURE 1.5 Mouvement orbital des particules en eau peu profonde [4].

La dispersion des vagues dépend de la longueur d'onde de celles-ci. Les vagues qui ont la longueur d'onde la plus grande se déplacent plus rapidement et quittent la zone de formation des vagues plus tôt. La dispersion engendre une ondulation sur l'eau qui donne naissance à la houle. Les ondes transmettent l'énergie, et non la masse d'eau à travers la surface de l'océan. Le comportement d'une onde dépend de la relation entre la taille de l'onde et la profondeur de l'eau à travers laquelle elle se déplace. Les vagues de vent se forment lorsque l'énergie est transférée du vent à l'eau [8].

1.1.5 Déferlement des vagues

Quatre types de vagues déferlantes se produisent sur la côte [9] [10]. Le type de déferlement d'une vague dépend des caractéristiques de celle-ci ainsi que de la pente du fond marin. Les quatre types sont les suivants :

- *Spilling waves* : Déferlement qui se produit lorsque le fond de l'océan a une pente douce. Au fur et à mesure que la vague s'approche du rivage, elle libère lentement de l'énergie et la crête se répand progressivement vers l'avant. Ces vagues mettent plus de temps à se briser que toute autre vague.

- *Plunging waves* : Déferlement qui a lieu quand la houle rencontre un fond océanique assez escarpé ou un fond marin avec des changements de profondeur soudains. En conséquence, la crête de la vague s'enroule et explose sur le creux. L'air sous la lèvre de la vague est comprimé.
- *Surging waves* : Déferlement qui surgit lorsque des houles de longue période arrivent sur les côtes avec des profils de plage escarpés. La base de la vague se déplace rapidement et ne permet pas à la crête d'évoluer. En conséquence, la vague ne se brise presque pas.
- *Collapsing waves* : Déferlement qui est qualifié comme un mélange de *surging wave* et *plunging wave*. La crête ne se brise jamais complètement et la face inférieure de la vague devient verticale et s'effondre.

La figure 1.6 illustre ces différents types de déferlement.

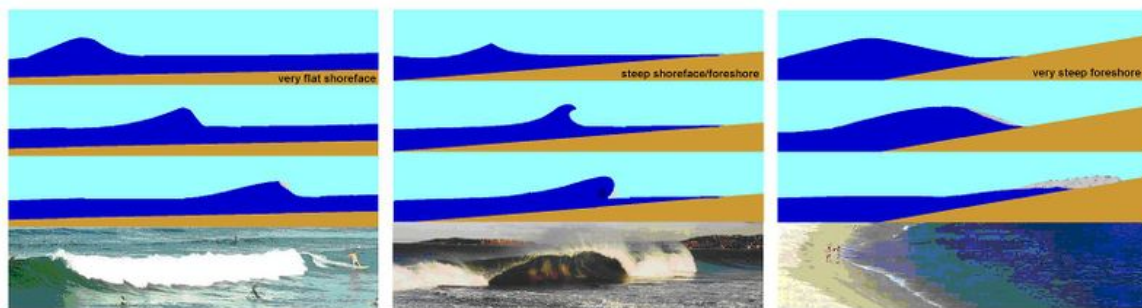


FIGURE 1.6 Types de déferlement des vagues. De gauche à droite : spilling, plunging and surging [11].

1.2 Les structures côtières

La principale raison de construire des structures côtières est de protéger les ports, les villes sur le littoral, et d'autres infrastructures. Ces structures protègent contre les effets des vagues comme l'érosion. Certaines de ces structures peuvent également être utilisées pour modifier les mouvements du courant et du sable ou encore pour rediriger les rivières.

Il existe diverses structures qui sont considérées et utilisées comme des structures côtières. Les plus communes sont les suivantes :

- Les épis (*groynes* en anglais),
- Les digues (*seawalls* en anglais),
- Les brise-lames (*breakwaters* en anglais),
- Les jetées (*jetties* en anglais).

1.2.1 Les différents types de structures

Dans cette section nous allons voir les différents types de structures évoqués en les décrivant et en donnant leurs avantages.

1.2.1.1 Epis

Un épi est une structure rigide de taille moyenne qui est construite perpendiculairement à la rive. Cette structure est réalisée à partir d'un rivage ou d'une rive. Elle a pour objectif d'interrompre l'écoulement de l'eau et limite le mouvement des sédiments. Un épi est généralement fait de bois, de béton ou de pierre. Les épis sont généralement placés en série.

Les épis océaniques s'étendent de l'estran supérieur ou de la plage jusqu'à l'eau. Un épi entier peut être sous l'eau, auquel cas il s'agit d'un épi submergé. Les épis sont souvent utilisés en combinaison avec des digues et d'autres éléments d'ingénierie côtière. Un épi crée progressivement une large zone de plage en piégeant les sédiments qui sont en suspension dans le courant océanique. Cela réduit l'érosion en réduisant la vitesse et la puissance des vagues qui frappent le rivage. C'est une barrière physique qui arrête le transport des sédiments en direction de la dérive littorale.

Les épis sont généralement rectilignes. Ils peuvent être perméables ou imperméables. Un épi mal conçu (trop long et inadapté aux caractéristiques uniques de la côte) peut en revanche accélérer l'érosion de la plage, qui reçoit peu ou pas de sable de la dérive littorale. La longueur et l'élévation d'un épi, ainsi que l'espacement entre les épis sont déterminés en

fonction de l'énergie locale des vagues et de la pente de la plage. Les épis trop longs ou trop hauts sont inefficaces car ils emprisonnent trop de sédiments. Les épis trop courts, trop bas ou trop perméables sont inefficaces car ils piègent trop peu de sédiments.

Les épis se distinguent par la façon dont ils sont construits, par leur effet sur l'écoulement des cours d'eau, par leur forme ainsi que par le fait qu'ils soient submergés ou non [12]. Ils sont classés en fonction de leur longueur, hauteur et perméabilité.

La figure 1.7 montre deux épis sur une plage où nous pouvons constater un mécanisme d'accrétion avant le premier épi et entre les deux épis, ainsi qu'un phénomène d'érosion après le deuxième épi.

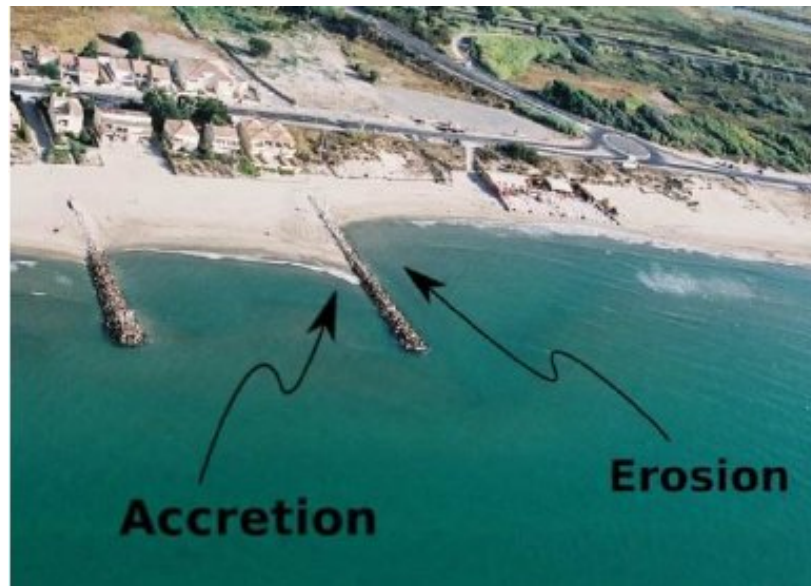


FIGURE 1.7 Exemple d'implantation d'épis où un mécanisme d'érosion/accrétion autour de l'ouvrage est observable [13].

1.2.1.2 Dignes

Une digue est une forme de défense côtière qui est construite là où la mer a un impact direct sur les reliefs de la côte. En effet, les digues sont généralement utilisées pour protéger le littoral contre l'érosion. Elles constituent aussi une méthode courante pour protéger les routes et les structures à risque. Elles peuvent également aider à atténuer les problèmes d'inondations. Ces grandes structures peuvent être construites avec une large gamme de matériaux de construction, dont les plus communs sont des blocs de pierre (enrochement), du granit ou du béton armé.

Une digue est construite directement sur le rivage. L'objectif d'une digue est d'éviter que les structures du littoral et les zones côtières soient influencées par les effets néfastes des vagues et des inondations causées par les tempêtes. Le but d'une digue est de protéger les zones d'habitation humaine, de conservation et de loisirs, de l'action des marées, des vagues ou des tsunamis [14]. C'est pour cela qu'elles sont destinées à résister et à dévier ou dissiper l'énergie des vagues qui les percutent.

Une digue influencera les modèles d'érosion environnants. La digue étant une caractéristique statique, elle entre en conflit avec la nature dynamique de la côte et entrave l'échange de sédiments entre la terre et la mer [15]. Les plages, là où une digue est construite, sont parfois perdues. Cet effet est causé lorsque les vagues sont si fortement déviées qu'elles créent un fort lavage à contre-courant qui entraîne le sable de la plage avec elle. Cet effet peut être atténué en créant une paroi qui comporte des rainures pour dissiper plus efficacement la force contre la paroi.

Comme elles sont directement placées sur le rivage, il est essentiel qu'elles soient bien construites afin de résister à toute la force des marées et que le matériau le plus approprié soit sélectionné pour l'emplacement et les fins principales prévues.

Il existe divers agencements et configurations qui peuvent être employés. Parmi ces différences, on peut noter les suivantes [16] :

- Digue à face incurvée (*curved face seawall*),
- Digue en escalier (*stepped face seawall*),
- Digue en décombres (*rubble mound seawall*).

La digue à face incurvée est conçue pour résister aux vagues puissantes. La digue en escalier est utilisée pour s'opposer ou résister aux actions modérées des vagues. La digue en décombres est plus facile à mettre en place et moins chère. Elle peut résister à l'action de vagues relativement fortes.

Si les digues sont correctement conçues en prenant en compte la situation spécifique du site, elles sont efficaces et permettent de préserver le bord de mer. Si elles sont suffisamment hautes et solides, elles peuvent protéger une zone d'arrière-plage contre les assauts des vagues de tempête. De manière générale, les digues dégradent les plages au fil du temps.

La figure 1.8 montre la photo d'une digue sur une plage.



FIGURE 1.8 Exemple d'une digue sur une plage [17].

1.2.1.3 Brise-lames

Contrairement à une digue qui est construite directement sur le rivage, un brise-lames est construit dans l'eau le long du rivage. Un brise-lames peut soit s'étendre dans la mer à partir du rivage, soit ne pas être attaché au rivage du tout, auquel cas il est qualifié de brise-lames détaché. Un brise-lames est généralement construit à partir de roches ou de béton, créant un mur ou même une barrière parallèle à la rive. Il est placé au large dans des eaux relativement peu profondes. Cette barrière brise les vagues avant que celles-ci n'atteignent le rivage, dissipant l'énergie de ces dernières. Une utilisation courante d'un brise-lames est de créer un petit port où les bateaux pourraient être amarrés tout en étant protégés du large.

Un brise-lames peut être submergé. Dans ce cas, il est construit pour être sous la surface de l'océan. Il est généralement placé dans le but de réduire les effets de l'érosion. Ceci est parfois même appelé récif artificiel.

Cependant, que le but soit de réduire directement l'érosion ou de fournir une zone protégée pour les bateaux, l'utilisation d'un brise-lames crée toujours des effets d'érosion

supplémentaires dans les zones environnantes. En effet, un brise-lames rompt l'écoulement naturel des sédiments, tant dans sa zone immédiate que le long du rivage environnant. Le brise-lames lui-même a tendance à agir comme un aimant à sédiments, mais peut également entraîner un appauvrissement important des zones qui l'entourent en sable et en sédiments. L'installation d'un brise-lames nécessite une surveillance attentive des sédiments dans les zones immédiates et environnantes.

Il existe deux types principaux de brise-lames :

- Les brise-lames en enrochement,
- Les brise-lames à caissons.

Les brise-lames en enrochement utilisent des vides structurels pour dissiper l'énergie des vagues. Ils sont constitués de piles de pierres plus ou moins triées en fonction de leur poids : des pierres plus petites pour le noyau et des pierres plus grosses comme couche de blindage protégeant le noyau des attaques des vagues. Les blocs de roches ou de béton à l'extérieur de la structure absorbent la majeure partie de l'énergie, tandis que les graviers ou les sables empêchent l'énergie des vagues de continuer à travers le noyau du brise-lames. Dans les eaux peu profondes, ces brise-lames sont généralement relativement peu coûteux. À mesure que la profondeur de l'eau augmente, les besoins en matériaux - et donc les coûts - augmentent considérablement [18].

Les brise-lames à caissons ont généralement des côtés verticaux et sont généralement érigés là où il est souhaitable d'accoster un ou plusieurs navires sur la face intérieure du brise-lames. Ils utilisent la masse du caisson et le remblai à l'intérieur pour résister aux forces de renversement appliquées par les vagues qui les frappent. Ils sont relativement coûteux à construire en eau peu profonde, mais dans les sites plus profonds, ils peuvent offrir une économie significative par rapport aux brise-lames en enrochement.

Un monticule de gravats supplémentaire est parfois placé devant la structure verticale afin d'absorber l'énergie des vagues et ainsi réduire la réflexion des vagues et la pression des vagues sur la paroi. Une telle conception offre une protection supplémentaire du côté de la mer et une paroi de quai du côté intérieur du brise-lames.

La figure 1.9 montre un brise-lame en reconstruction. Sur cette photo, nous voyons bien que la structure s'étend dans l'eau, et que celle-ci est présente des deux côtés.



FIGURE 1.9 Exemple d'un brise-lame en reconstruction [19].

1.2.1.4 Jetées

Les jetées sont généralement construites avec des matériaux tels que du bois, de la terre, de la pierre ou du béton. Ces structures sont construites aux estuaires de rivières ou à l'entrée des ports. Elles sont étendues dans les eaux plus profondes pour empêcher la formation de bancs de sable et limiter les courants.

Les jetées sont des structures longues et étroites qui protègent le littoral des courants et des marées. En effet, elles protègent le rivage d'un plan d'eau en agissant comme une barrière contre l'érosion causée par les courants, les marées et les vagues. De plus, les jetées peuvent aussi être utilisées pour relier la terre à des eaux profondes plus éloignées du rivage afin d'accoster les navires et de décharger la cargaison. Elles offrent généralement un accès sûr aux zones côtières.

La différence majeure entre les jetées et les épis est l'étendue de la structure. En effet, une jetée est généralement plus longue et plus étroite qu'un épi. De plus, les jetées ne sont pas placées en série, contrairement aux épis.

La figure 1.10 montre des photos aériennes d'une longue jetée à Huelva en Espagne.



FIGURE 1.10 Exemple d'une longue jetée [20].

1.2.2 Effet de la houle sur les structures

Lorsque qu'une vague rencontre une structure côtière comme une digue ou un brise-lames, l'énergie qu'elle transporte se fractionne à la rencontre de l'obstacle. En effet, une partie de l'énergie est dissipée par turbulence lors du déferlement. De plus, la rugosité de la paroi percutée crée des frottements qui dissipent l'énergie. Si l'obstacle est perméable, une partie de l'énergie passe à travers, sinon l'énergie est transmise au-dessus. Enfin, une partie de l'énergie est réfléchiée par l'obstacle.

Ces structures arrêtent la houle en encaissant l'énergie de celle-ci, ayant pour effet de la supprimer. Bien que ces ouvrages soient conçus pour résister à l'action des vagues, le temps et les impacts de vagues répétés finissent par les user et les endommager.

En effet, lors de grosses tempêtes ces structures peuvent être sévèrement touchées. Les cas les plus extrêmes mais aussi les plus rares sont la rupture de l'ouvrage ou le basculement de celui-ci. Par exemple, la figure 1.11 montre un brise-lames détruit lors d'un tsunami. Des conséquences plus classiques et plus régulières sont des arrachements de parties de la structure. Ces arrachements sont problématiques car s'ils ne sont pas régulièrement réparés, l'eau continue de creuser et de s'infiltrer, fragilisant la structure.



FIGURE 1.11 Brise-lames détruit à Takenoura, Japon, lors du tsunami en 2011 [21].

Il est donc important d'entretenir ces ouvrages. Cependant, cela peut s'avérer très complexe à cause de l'environnement hostile dans lequel ils sont souvent situés. De ce fait, maintenir ces structures demande de gros moyens techniques et coûte relativement cher.

1.3 Objectifs

Il y a donc un intérêt à surveiller au mieux les structures côtières, afin de pouvoir agir avant que les dégâts ne soient trop importants. C'est dans cette optique que nous avons souhaité étudier les impacts de vagues sur les structures côtières. Pour atteindre cet objectif, trois équipes de recherches se sont associées :

- SIAME IVS, qui étudie les interactions entre les vagues et les structures,
- SIAME PHT, qui travaille notamment sur le développement de dispositifs de mesure HT,
- LIUPPA, qui est un laboratoire informatique.

Cette collaboration a permis de déployer une station de mesures sur une structure côtière, avec l'aide du Conseil départemental des Pyrénées-Atlantiques. Cette station a pour objectif de récolter des données en continu afin de pouvoir superviser la structure et de mieux comprendre les causes et les effets de la houle sur cet ouvrage.

La station permet donc de récolter des données en masse. Il a ensuite fallu mettre en place des algorithmes permettant de traiter toutes ces données. Le but étant d'essayer de trouver et d'extraire de l'information dans ces données, permettant à terme une meilleure supervision de la structure. Nous retrouvons donc ici les 4V du Big Data :

- Volume : un grand nombre de données est enregistré et traité,
- Vitesse : des données sont enregistrées à haute fréquence et à chaque heure,
- Variété : des données provenant d'autres stations de mesures sont utilisées,
- Valeur : extraire de l'information utile de ces données.

Chapitre 2

Etat de l'art

Dans ce chapitre, nous allons faire un état de l'art. Celui-ci se décomposera en deux grandes parties :

1. Pression d'impact des vagues,
2. Apprentissage automatique.

Dans la première partie, nous allons resenser les études existantes qui nous ont permis de réfléchir et d'établir un cheminement jusqu'aux résultats que nous avons obtenus.

Dans la deuxième partie, nous allons faire une présentation de l'exploration de données, de l'apprentissage supervisé et de l'apprentissage non-supervisé, techniques permettant de valoriser une masse de données en y extrayant de l'information.

2.1 Pression d'impact des vagues

Les impacts des vagues ont été identifiés très tôt comme l'une des principales sources causant des dommages aux structures côtières [22]. C'est pour cette raison que la force des vagues sur ce type de structures a commencé à être étudiée avant le début du XXème siècle. Ce phénomène a été étudié à la fois en laboratoire et sur le terrain. Les premières tentatives de mesure de pression d'impact de vagues sur une structure réelle ont été réalisées au XIXème siècle par Thomas Stevenson. Celles-ci ont échoué en raison du manque de fiabilité de l'équipement de l'époque. Ce n'est que dans les années 1930, lorsque des instruments électroniques plus modernes ont été disponibles, que de meilleures mesures ont pu être obtenues. Cependant, le volume de données à grande échelle disponibles est encore aujourd'hui remarquablement faible. La raison principale est la difficulté à effectuer des mesures de précision dans un environnement extrêmement hostile comme celui des océans [23].

Divers modèles ont été développés grâce aux données disponibles obtenues sur site réel. L'objectif de ces modèles est d'aider les concepteurs à évaluer les pressions d'impact maximales que les vagues déferlantes peuvent imposer à une structure côtière. Il s'agit notamment de la formule du British Standard 6349 [24], développée par Blackmore et Hewson [25], ainsi que la formule du Centre de recherches en génie côtier (CERC pour *Coastal Engineering Research Centre*) [26] qui est basée sur l'étalonnage du modèle de Bagnold [27]. Ces formules ont été examinées par Muller et Whittaker [28] qui ont conclu qu'aucune des deux n'incluait tous les paramètres pertinents et qu'elles ne fonctionnaient pas avec leurs propres mesures sur le terrain.

Alors qu'il n'y a eu que quelques études sur la pression d'impact des vagues qui ont été réalisées à grande échelle, il en existe plusieurs qui ont été réalisées en laboratoire. Le problème de ces études en laboratoire, est qu'il est difficile d'utiliser ces résultats pour fournir des estimations réalistes des conditions sur le terrain. En effet, il y a une problématique de mise à l'échelle qui s'avère être complexe.

Dans cette section, nous allons décrire les principales études sur lesquelles nous nous sommes basées pour guider nos travaux.

2.1.1 Etudes in-situ

2.1.1.1 De Rouville

Les résultats obtenus par De Rouville *et al.* [29] sont le fruit d'un long travail, de 1927 jusqu'à 1938. Avant cela, la pression des vagues n'avait a priori jamais été enregistrée de manière continue. En effet, les études antérieures donnaient la valeur moyenne ou la valeur maximum de la pression pendant un temps déterminé. Afin de pouvoir mieux analyser le phénomène, ils ont voulu enregistrer en continu afin de pouvoir tracer la courbe représentant les variations en fonction du temps, plutôt que de se contenter d'une seule valeur.

Cette étude repose sur l'analyse de résultats provenant de plusieurs sites. En effet, ils ont travaillé sur des structures côtières situées à différents endroits :

- à Dieppe,
- au Havre,
- à Alger.

A partir de 1927, ils ont entrepris la mise en place d'un premier appareil hydraulique permettant l'obtention de telles courbes. Cela permettait donc de suivre les variations des pressions qui peuvent se produire, en particulier lorsqu'il y a déferlement. Plusieurs appareils de différents types ont été testés et déployés.

Le premier type correspond à des appareils mécaniques. Ces appareils sont trop peu sensibles au-delà d'une certaine fréquence. Ils ne sont donc pas aptes à reproduire les variations très brusques et rapides des efforts appliqués par le choc des vagues. Un deuxième type d'appareils sont les appareils à quartz. Ces appareils ont une perte de sensibilité dans les fréquences très basses, c'est-à-dire lors des efforts très lentement variables. Enfin, le dernier type correspond aux appareils électriques.

Lors de leurs expériences, ils ont essayé différents appareils. Comme appareils mécaniques, ils ont utilisé un manographe à transmission hydraulique. Celui-ci leur a permis d'étudier les efforts sans déferlement. Ils ont également utilisé un manographe différentiel. Cependant, ils souhaitaient pouvoir effectuer des enregistrements à des fréquences plus élevées, afin de pouvoir observer les variations brusques.

Pensant que les appareils utilisant des cristaux de quartz piézoélectrique étaient trop fragiles pour l'environnement extrême d'une digue, ils se sont tournés vers des appareils électriques qui permettraient d'analyser les variations des efforts en cas de déferlement. Ces appareils ont donc été testés en laboratoire, avec succès. Cependant, la mise en place sur site réel a posé problème.

Entre temps, ils se sont rendu compte que les appareils à quartz n'étaient pas si fragiles et convenaient donc à leurs besoins. Ils ont finalement installé ce type de capteurs piézoélectriques afin d'enregistrer les variations de pressions lors de déferlements. En effet, ces appareils ont une fréquence propre élevée. C'est pour cela qu'ils ont été utilisés à Dieppe et au Havre, permettant d'enregistrer les variations très rapides de pression dans les chocs des lames déferlantes. Ils ont également été utilisés à Dieppe et Alger afin d'enregistrer la pression lentement variable du clapotis.

A Dieppe, l'installation compte 4 manographes, dont 3 sont fixes. Ils sont placés verticalement, espacés d'environ 1 mètre à chaque fois. Le manographe inférieur se trouve à une hauteur de 0.37 mètre. Le manographe intermédiaire est à une hauteur de 1.34 mètre. Enfin, le manographe supérieur se trouve à une hauteur de 2.28 mètres. Ces 3 manographes ont pour objectif d'enregistrer la pression des lames déferlantes. Ils peuvent admettre une pression de 100 mètres d'eau. 0.1 bar valant environ 1.02 mètre d'eau, ces capteurs étaient capables d'enregistrer des pressions de 9.8 bars environ. Le quatrième manographe de cette installation est un manographe fixé sur un panneau mobile, pouvant bouger sur un axe vertical de la jetée ouest du port de Dieppe qui est plane et verticale. Il permettait d'enregistrer les pressions à une hauteur spécifique du mur, y compris jusqu'au fond de l'eau, à -4 mètres.

Au Havre, l'installation était relativement identique à celle de Dieppe, sauf qu'il y avait 5 manographes fixes. Ces capteurs n'étaient pas seulement placés de manière verticale, mais aussi horizontale, de part et d'autre de la région la plus vivement frappée par les lames.

Ces installations nécessitaient un poste de réception, dans lequel se trouvaient un amplificateur, un oscillographe, des batteries d'accumulateurs et un chauffage électrique permettant de garder la température au-dessus de 18°. Un groupe électrogène avait pour objectif de recharger les accumulateurs. Une salle à l'abri des vagues était nécessaire afin de contenir tous ces instruments. Par exemple à Dieppe, cette salle abritée se trouve à 300m des capteurs. Une telle installation demandait de très gros moyens et était complexe à mettre en place.

Un observateur, placé à un endroit stratégique, avait pour mission de démarrer manuellement, au moment voulu, l'oscillographe. En parallèle, il devait apprécier l'amplitude et la célérité de la houle par observation grâce à une lunette stadimétrique.

Ces expériences ont permis de dégager des résultats sur différents phénomènes, comme sur la houle à mouvement orbitaire. En effet, ces instrumentations ont permis de comparer les résultats des expériences avec les formules grâce aux enregistrements de hauteur et de

pression au parement d'une digue verticale. Les résultats qui nous intéressent le plus dans notre cas sont ceux obtenus sur les conditions du déferlement.

A Dieppe, des enregistrements cinématographiques accompagnaient les enregistrements manographiques de pressions dues au choc de lames déferlantes et des vitesses d'embruns projetés. Pour cela, un appareil de prise de vue a été utilisé. Celui-ci avait une fréquence de 24 images par seconde.

Un contrôleur placé prêt de la caméra centralisait les circuits électriques permettant de commander la mise en route de tous les appareils (caméra et oscillographe) lorsqu'il souhaitait enregistrer une vague qu'il avait déterminé comme pertinente.

Ces enregistrements et prises de vues ont permis de déterminer que des lames d'apparence identique déferlaient très inégalement. Cela provenait de différents processus :

- Certaines étaient freinées par l'écoulement des précédentes,
- D'autres frappaient le mur obliquement ou loin des capteurs,
- Quelques-unes seulement s'enroulaient au pied du mur, enfermaient un bon volume d'air et claquaient contre la paroi à l'endroit des capteurs.

Ils ont considéré que la proportion d'enregistrement intéressant était inférieure à 5%. La pression maximum enregistrée a été de 70 tonnes par mètre carré, ce qui vaut environ 6.865 bars. Cette pression a été appliquée sur le capteur inférieur. La pression maximum sur le capteur intermédiaire a été de 30 tonnes par mètre carré. Enfin, le capteur supérieur a souvent été sans choc. Seulement 2 fois le manographe intermédiaire a enregistré une pression plus forte que le manographe inférieur. Cela permet tout de même de déterminer que le choc principal s'exerce à des hauteurs variables. La mise en charge de 0 à 70 tonnes par mètre carré ne dure parfois que 0.005 sec, ce qui est extrêmement rapide.

Les films montrent des petites « fusées » d'embruns qui partent à des vitesses surprenantes (77 m/s). De grands panaches sont lancés en l'air à des vitesses moindres et pourtant encore considérables (35 m/s).

Ils ont donc conclu que les plus grosses lames ne donnent pas nécessairement les chocs les plus violents. En effet, ils montrent que l'air est comprimé dans les lames lors du déferlement et qu'il s'échappe dans la direction de moindre résistance en projetant à très grande vitesse une fraction de son enveloppe d'eau. Ces jets peuvent provoquer de fortes pressions. Une masse d'eau notable projetée par une quantité importante d'air comprimé est susceptible de causer de gros dégâts bien que généralement très localisés.

2.1.1.2 Bird et Bullock

Bird *et al.* [23] en 1998 ont mis au point un équipement conçu pour faciliter la mesure simultanée des pressions d'impact des vagues, des niveaux d'aération et du profil des vagues incidentes.

Ces capteurs permettent de mesurer en simultané la pression des impacts des vagues et le niveau d'aération de l'eau de mer lors du choc. Ils ont été développés et déployés sur un site d'essai exposé aux vagues de tempête, à Alderney.

Le transducteur de pression d'impact dispose d'une surface de détection de 18 mm de diamètre, est robuste, non corrodable, et a une fréquence naturelle de 2,5 kHz. Le transducteur d'aération mesure indirectement la fraction d'air dans l'eau de mer dans une région autour du capteur de pression grâce aux variations de conductivité électrique du mélange air-eau. Un système d'acquisition de données complémentaire a été développé sur la base d'un ordinateur et d'un dispositif de stockage de données sur bande audio numérique. Sept capteurs de pression et d'aération ainsi que des instruments auxiliaires comprenant un accéléromètre, une caméra vidéo et un système d'enregistrement des vagues ont été déployés.

Par la suite, Bullock *et al.* [30] ont réalisé des mesures sur le terrain durant 8 mois, là aussi à Alderney. Ils ont utilisés 8 capteurs conçus par Bird *et al.*, décrit précédemment. Cette installation leur a permis d'effectuer des mesures de pression d'impact de vagues à Alderney. Dans cette étude, la pression maximale enregistrée n'a pas dépassé celle d'une précédente étude également menée par Bullock *et al.* [31], où la pression maximale enregistrée était de 3.37 bars à 6.2 mètres de hauteur avec des valeurs plus faibles à des niveaux plus élevés. Des hauteurs de vagues significatives allant jusqu'à 3.1 mètres ont été enregistrées.

2.1.2 Etudes en laboratoire

2.1.2.1 Hofland

L'étude de Hofland *et al.* [32] en 2010 a pour objectif de réaliser et d'analyser des mesures de pression d'impact de vagues à grande échelle sur une paroi verticale. Pour ce faire, un mur de 9 mètres de haut a été construit, composé de deux couches de béton de 0,75 m d'épaisseur soutenues par trois supports en acier. Un panneau de 2 mètres de haut, contenant l'instrumentation, a été incorporé dans la face avant du mur en béton. La disposition des éléments en béton de la paroi avant pouvait être modifiée pour permettre des tests avec différentes positions de capteur. Le mur a été placé à une distance de 145,2 m du générateur de vagues.

Pour permettre l'étude de différents types d'impacts de vagues déferlantes, un signal de vague focalisée a été développé, qui produit une vague qui se propage sur un lit d'eau plat et se brise à un endroit désigné. Un algorithme a été conçu pour contrôler la génération des signaux de vagues. Le point de rupture de la vague pouvait être déplacé explicitement. Ainsi, le type d'impact des vagues peut être modifié sans modifier la configuration ou le niveau d'eau.

Au total, 23 capteurs de pression ont été mis en place, avec une fréquence d'échantillonnage de 25 ou 50 kHz. La distance minimum entre deux capteurs de pression était de 12 cm, à l'endroit de la zone d'impact. Plus loin de la zone d'impact, l'espacement des capteurs variait entre 0,2 m et 1 m. Les capteurs de pression miniatures qui ont été installés dans la région où les pressions les plus importantes se produisent, ont été conçus sur mesure pour ces tests. De plus, une isolation thermique a été conçue pour éliminer les chocs thermiques. Ils ont un diaphragme de 3,6 mm de diamètre.

Au total, 137 impacts ont été mesurés avec des profondeurs d'eau variant de $h = 2,5$ à $h = 4,25$ m. La disposition des éléments en béton de la paroi avant a été modifiée pour permettre des tests avec différentes profondeurs d'eau / positions de capteurs.

Les formules disponibles pour estimer les charges sur les murs verticaux soumis à des vagues déferlantes ont donné des résultats très différents de ceux observés sur le mur pendant l'étude.

Plusieurs types d'impacts ont été identifiés précédemment [33], [34], [35]. Il a été possible d'obtenir plusieurs types d'impact qui ressemblent à ceux mentionnés dans la littérature lors des essais en changeant le point focal de la vague. Ces quatre types d'impacts sont :

- *aerated impact* : la vague a été brisée et une masse d'eau aérée frappe le mur, cet impact se produit pour les vagues qui se brisent avant d'avoir atteint le mur,
- *air pocket impact* : au moment de l'impact, la crête de la vague frappe le mur, de telle manière qu'une bulle d'air se trouve piégée entre le mur et la vague,
- *flip through impact* : le creux de la vague et la crête de la vague se rencontrent au moment de l'impact sur le mur, de sorte que (presque) tout l'air a été expulsé de la poche d'air au moment de l'impact [36],
- *slosh impact* : la montée de la vague est plus élevée que la crête de la vague, de sorte que la crête de la vague frappe la couche d'eau au lieu du mur. Ce type d'impact se produit pour les vagues qui se briseraient loin derrière le mur.

La figure 2.1 schématise ces quatre types d'impacts. Cette figure permet de mieux imaginer et comprendre le moment du déferlement par rapport à une paroi ainsi que de visualiser la poche d'air piégée lors du *air pocket impact* et du *flip through impact*.

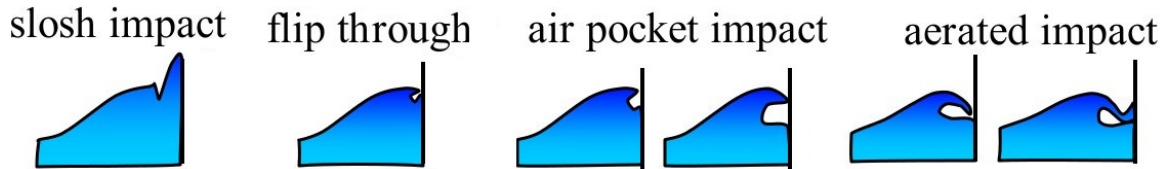


FIGURE 2.1 Types d'impact des vagues en fonction du déferlement [32].

Lors de leur étude, ils ont constaté que les impacts *flip through* sont ceux qui ont créé les pressions les plus importantes. Seuls 8 impacts avaient des pressions supérieures à 1 MPa (10 bars). La plupart des pics de pression les plus élevés avaient une durée inférieure à une milliseconde. Cependant, certains avaient une durée d'environ 10 ms. La plus grande pression d'impact mesurée a été de 2,7 MPa, avec une durée très courte (<0,5 ms). Cela est comparable au pic de pression de 3,5 MPa mesurée dans [35]. Les pics de pression très élevés sont très localisés sur l'axe vertical. Les impacts des *air pocket* génèrent des pressions comprises entre 0,2 et 0,5 MPa, avec une tendance à des pressions plus élevées pour les poches d'air plus petites. Les *aerated impact* et les *slosh impact* montrent des pressions quasi-statiques, de l'ordre de la hauteur des vagues.

Hofland *et al.* ont étudié la pression d'impact des vagues à grande échelle dans le canal de Delta. Des vagues d'une hauteur maximale de 2,6 mètres ont été générées permettant d'observer plusieurs types d'impacts sur un mur vertical. La forme de la vague et la distribution de la pression ont été mesurées à grande échelle.

Ils ont constaté une grande variabilité des forces et des pressions. La variabilité est la plus grande pour les impacts *flip through*. Ils ont observé que les impacts *flip through* ne se produisaient que de manière très localisée, ce qui les rend rares et difficiles à observer. Les impacts *flip through* ont généré les pics de pressions et les pics de forces les plus importants. Les impacts de type *air pocket* ont également généré de grandes pressions et de grandes forces.

La figure 2.2 montre la génération de vague dans le canal de Delta.



FIGURE 2.2 Exemple de génération de vague dans le canal de Delta [32].

2.1.2.2 Kaminski

Kaminski *et al.* [37] en 2009 ont également étudié les impacts de vagues à grande échelle. Pour cela, des vagues ont été générées en canal, où là aussi elles percutaient un mur hautement instrumentalisé permettant l'enregistrement de la pression d'impact des vagues sur le mur.

Ce qui nous intéresse dans cette étude, ce sont les signaux des différents types d'impacts sur un axe vertical. La figure 2.3 nous permet de visualiser le signal de pression d'impact pour chaque type d'impact, à différentes hauteurs du mur.

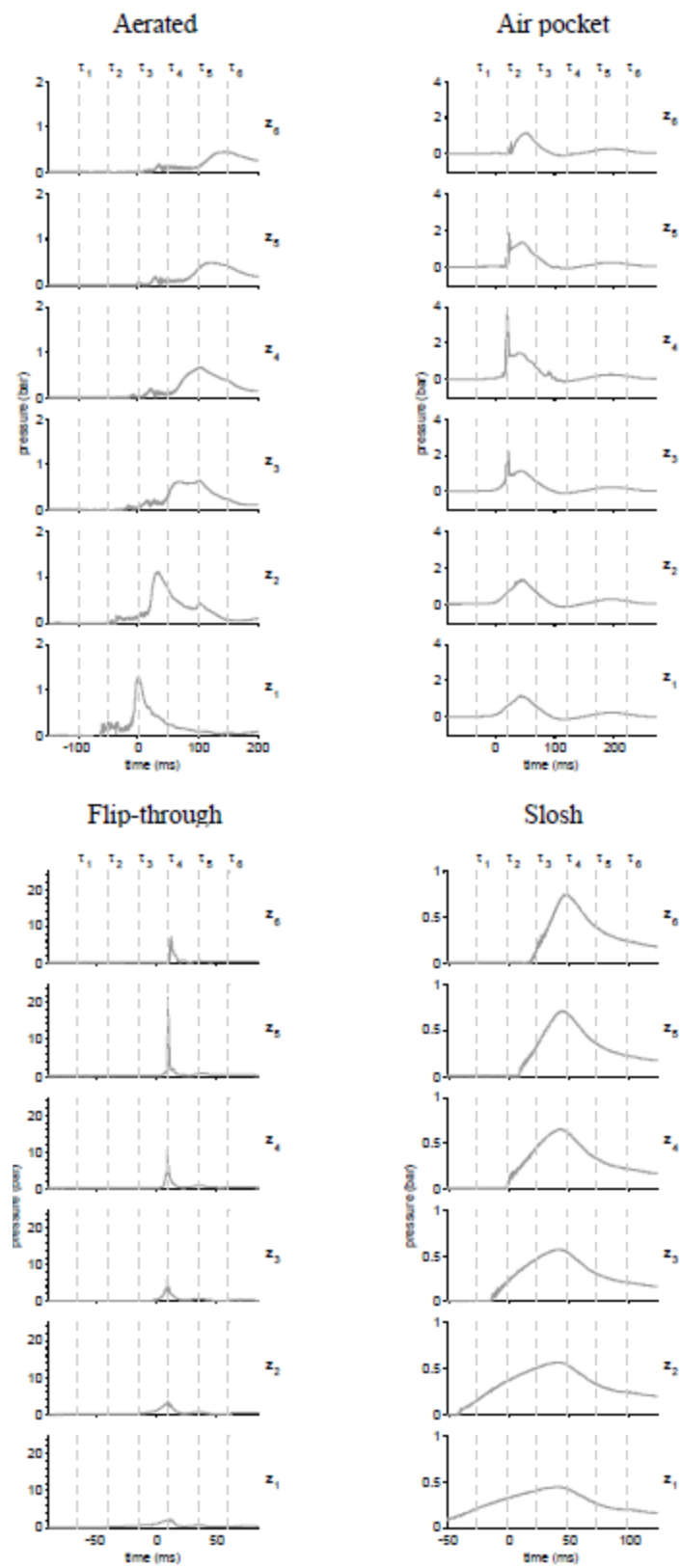


FIGURE 2.3 Signal d'impact pour chaque type d'impact à différentes hauteurs d'une paroi [37].

Cette figure est très intéressante car elle montre que chaque type d'impact a ses propres caractéristiques et donc l'allure d'un signal d'impact permet de déterminer visuellement de quel type d'impact ils'agit. En effet, nous pouvons observer des différences d'échelle et de durée. Le *flip through* génère la pression la plus élevée (environ 20 bars) sur une durée très faible (environ 50 ms). Le deuxième type le plus puissant est le *air pocket* dont la pression monte jusqu'à 4 bars, avec une durée d'impact de 200 ms. L'*areated* est moins puissant, générant une pression de 2 bars et ayant une durée de 100 ms. Enfin, le *slosh* dure également 100 ms mais a une faible pression, de seulement 1 bar.

2.1.3 Bilan

Nous pouvons extraire plusieurs informations importantes de cette partie. La première est qu'il y a peu d'études sur les impacts de vagues sur les structures côtières qui sont réalisées sur le terrain à cause de l'environnement hostile dans lequel ces ouvrages se trouvent. C'est pour cela que la majorité des études sont réalisées en laboratoire, où il est plus simple d'observer et répéter des expériences.

Ensuite, nous avons vu qu'il y a quatre principaux types d'impacts, qui génèrent des pics de pression différents et de durées différentes. La figure 2.1 montre ces différents types d'impacts.

En lien avec ce dernier point, nous avons pu voir que chaque type d'impact a un signal bien caractéristique, permettant de les différencier. Cela est observable sur la figure 2.3.

Enfin, en comparant les résultats obtenus entre les études en laboratoire et les études sur le terrain, nous pouvons observer que les valeurs des pressions d'impacts ne sont pas du même ordre de grandeur. En effet, en laboratoire des pics de pression jusqu'à 27 bars ont été enregistrés [32], alors que sur le terrain, le pic de pression le plus important enregistré jusqu'à aujourd'hui a une valeur de 6.9 bars environ [38]. Il y a donc une grosse différence entre ces résultats, ce qui a pour conséquence de complexifier l'utilisation des résultats obtenus en laboratoire pour l'interprétation de résultats obtenus sur le terrain.

Pour finir, comme il est suggéré dans [37] et affirmé dans [27], même dans des conditions de vagues fixes, pour un point donné sur une paroi, le pic de pression change de manière imprévisible entre des impacts de vagues apparemment identiques.

2.2 Apprentissage

Le *Data mining* passe souvent par une phase d'apprentissage automatique. C'est cette dernière qui permet d'extraire des modèles et des connaissances de manière automatique dans les données. Cet apprentissage peut être réalisé grâce à deux approches différentes du *machine learning*, l'apprentissage supervisé et l'apprentissage non supervisé. Le choix entre l'un et l'autre dépend directement des connaissances déjà disponibles sur le jeu de données utilisé et les informations qu'il faut en extraire.

2.2.1 Exploration de données

Les progrès de la technologie ont permis d'améliorer l'acquisition de données numériques et le stockage des données. Cela a inévitablement entraîné une augmentation de la taille des bases de données, qui peuvent devenir énormes. De ce fait, extraire de l'information de cette masse de données peut s'avérer complexe. Cependant, il y a un intérêt à exploiter au maximum ces données afin d'en extraire le plus d'informations possible. Ce processus est connu sous le nom d'exploration de données, ou *data mining*.

2.2.1.1 Qu'est-ce que l'exploration de données ?

L'exploration de données consiste à analyser des ensembles de données volumineuses (le premier V du Big Data : Volume) afin d'y trouver des relations et de présenter les données d'une nouvelle manière, à la fois compréhensible et utile (le quatrième V du Big Data : Valeur). Les relations et les informations dérivées du processus de *data mining* sur des données sont souvent appelées modèles. Les modèles incluent les équations linéaires, les règles, les clusters, les graphiques, les structures arborescentes, les modèles récurrents dans les séries chronologiques, etc. [39].

Le *data mining* consiste donc à rechercher des modèles dans un ensemble de données. Ces modèles permettent par la suite de faire de la classification, de la prévision, de la recommandation, etc.

2.2.1.2 Ensemble de données

Un ensemble de données est un ensemble de mesures collectées dans un environnement ou dans un processus. Dans les cas classiques, un ensemble de données représente un ensemble de n objets (aussi appelés individus ou entités). Chaque objet est représenté par un vecteur v qui contient une série de mesures propres à cet objet. Ces mesures peuvent être appelées variables, caractéristiques ou encore attributs. Tous les objets de l'ensemble de

données ont les mêmes mesures. L'ensemble de données peut donc se représenter comme une matrice $n \times v$, où chaque ligne correspond à un objet et chaque colonne correspond à un type de mesure.

La figure 2.4 montre un extrait d'un ensemble de données provenant de la station de la que nous avons mis en place, que nous présenterons en détail dans la partie 3. Ici, chaque ligne représente une acquisition de données et chaque colonne représente une caractéristique (mesure) de chaque acquisition. Les acquisitions correspondent donc aux objets et les colonnes aux caractéristiques de ces objets.

file	date	P1/10	Int1/10	mean	nbImpacts	Pmax	Tpmax	Intpmax	Patm
20190121-040001-10min-712Imp-rate21.raww	21/01/2019 04:00:01	1,136736137	0,41152356	0,015488193	58	4,472591656	3,704101563	0,262174157	1,04027087
20190130-010001-10min-41408Imp-rate21.raww	30/01/2019 01:00:01	1,64150489	2,230188422	0,085510395	47	2,545038224	2,434082031	0,252588227	1,03447653
20190127-210001-10min-64828Imp-rate21.raww	27/01/2019 21:00:01	1,665478097	2,833944793	0,101639086	43	1,994211959	9,169042969	3,14400128	1,04091582
20190123-060001-10min-42529Imp-rate21.raww	23/01/2019 06:00:01	1,150738994	2,385078847	0,077571346	46	1,937905567	17,32783203	2,733403581	1,02310907
20180615-170001-10min-60307Imp-rate21.raww	15/06/2018 17:00:01	0,54073089	0,413108793	0,018278646	58	1,778764814	7,359765625	0,227737857	1,06922139
20190127-220001-10min-48891Imp-rate21.raww	27/01/2019 22:00:01	1,553784299	2,438833244	0,087407917	45	1,770025867	7,097265625	2,276905181	1,0368408
20190219-020001-10min-25787Imp-rate21.raww	19/02/2019 02:00:01	0,766489789	1,126662785	0,018417669	34	1,700237302	3,424804688	0,219872464	1,06378132

FIGURE 2.4 Extrait d'un ensemble de données.

2.2.1.3 Knowledge Discovery in Databases

Il ne faut pas confondre *data mining* et "découverte de connaissances dans les bases de données", abrégé en KDD pour *Knowledge Discovery in Databases* en anglais. KDD qualifie le processus général de la découverte de connaissances à partir de données. Le *data mining* ne correspond donc qu'à une partie, certes centrale, du processus KDD [40].

Bien qu'il existe de nombreuses définitions du processus KDD ([41] [42] [43]), la plupart s'accordent sur les points essentiels. Fayyad et al. définissent le KDD comme un processus interactif et itératif. En effet, ils ont déterminé neuf étapes principales :

1. Identifier l'objectif du processus et rassembler les connaissances préalables nécessaires sur le domaine d'application.
2. Choisir un ensemble de données appropriées afin d'en extraire les connaissances souhaitées.
3. Prétraiter les données. Cela inclut la suppression du bruit introduit dans les données et/ou la suppression des données aberrantes ou erronées.
4. Réduire les données en un format pertinent et représentable. Cela peut passer par la suppression de paramètres inutiles pour l'objectif de la tâche.
5. Choisir une approche pour l'exploration des données en fonction de l'objectif défini lors du processus KDD.
6. Choisir l'algorithme le plus adapté pour répondre à l'objectif.

7. Appliquer l'algorithme sélectionné aux données prétraitées. L'algorithme permet la recherche de connaissance dans l'ensemble de données.
8. Interpréter les modèles dégagés par l'algorithme. Après cette étape, il y est possible d'ajuster le processus KDD en revenant à l'une des étapes précédentes.
9. Utiliser les résultats pour d'autres objectifs.

Comme indiqué à l'étape 8, après l'interprétation des résultats, le processus KDD peut être amené à itérer plusieurs fois afin d'ajuster différents paramètres du processus. En effet, il est possible d'avoir à modifier divers paramètres, tels que les données sélectionnées, le prétraitement appliqué, l'algorithme choisi, etc. Cela revient donc à revenir à une étape précédente et à reprendre le processus à partir de cette étape.


2.2.2 Standardisation des entrées

En *Machine Learning*, la standardisation des entrées est une étapes très importante, bien que simple à réaliser. Il s'agit de mettre toutes les variables à la même échelle (le troisième V du Big Data : Variété), afin que les variables dont les valeurs sont les plus grandes ne soient pas plus déterminantes que les variables dont les valeurs sont les plus petites. Par exemple, si nous travaillons avec l'âge de personnes et leur salaire annuel comme *features* (caractéristiques), l'échelle entre les deux valeurs est disproportionnée. En effet, si l'on prend par exemple une personne de 30 ans qui a un salaire annuel de 30000 €, il y a un facteur 1000 entre les deux valeurs, et donc, en l'état, le salaire sera beaucoup plus important dans le modèle d'apprentissage.

De plus, pour certains algorithmes comme les réseaux de neurones, la convergence est souvent plus rapide si les entrées sont à la même échelle et que la moyenne de chaque variable d'entrée sur l'ensemble d'apprentissage est proche de zéro [44].

Une façon simple de réaliser la mise à l'échelle est de retirer la moyenne sur chaque donnée d'entrée et de diviser par l'écart-type. Cela permet d'avoir une moyenne très proche de zéro et un écart-type de un. Bien entendu, il existe plusieurs méthodes de mise à l'échelle. La figure 2.5 montre un jeu de données pour cinq individus, où les entrées salaire annuel et âge sont standardisées afin d'être à la même échelle.

N° individus	Âge	Salaire
1	45	35000
2	30	30000
3	23	32000
4	48	27000
5	37	42000



N° individus	Âge	Salaire
1	0,40546132	0,15738692
2	-0,318576751	-0,27979897
3	-0,656461184	-0,10492461
4	0,550268934	-0,5421105
5	0,019307682	0,76944716

FIGURE 2.5 Standardisation des features.

2.2.3 Apprentissage supervisé

Cette approche permet d'extraire des connaissances sur un ensemble de données déjà labélisées. Ces données sont appelées les exemples. Chaque exemple est constitué d'un *input* qui correspond à un vecteur de données (les *features*) le caractérisant, et d'un *output* qui représente la sortie associée. Grâce à ces exemples, le modèle doit déduire une fonction de mappage (ou mise en correspondance) afin de pouvoir associer correctement à un *input*, l'*output* correct. Pour ce faire, le modèle va itérer sur tous les exemples qui lui sont fournies en essayant de prévoir l'*output* associé. En cas d'erreur, le modèle va s'ajuster pour se corriger. Cette approche permet donc d'entraîner le modèle sur un ensemble de données connues, ce qui permet de l'ajuster le plus précisément possible. Une fois bien entraîné, ce modèle est utilisé avec des données dont le résultat est inconnu, ce qui permet d'obtenir le résultat de ces dernières.

En général, l'ensemble de données d'exemples utilisé pour entraîner un modèle est séparé en deux :

- L'ensemble d'entraînement,
- L'ensemble de validation.

L'ensemble d'entraînement est utilisé pour entraîner et corriger le modèle afin que ce dernier s'ajuste le plus précisément possible. L'ensemble de validation quant à lui est utilisé après la phase d'entraînement, afin de vérifier si le modèle a effectivement bien appris. En effet, il est préférable d'avoir des données de validation que diffèrent des données d'entraînement car il est important de contrôler le modèle sur des données qu'il n'a jamais vues.

2.2.3.1 Réseaux de neurones

Un réseau de neurones artificiels est un système inspiré du fonctionnement du cerveau et des structures neuronales qui s'y effectuent. Le cerveau contient un nombre important de neurones. Un neurone est une cellule qui constitue la base du système nerveux qui traite les

signaux électriques. Chaque neurone est considéré comme une entité autonome qui interprète les informations qui lui parviennent et en déduit une information qu'il transmet à d'autres neurones. Pour expliquer les réseaux de neurones artificiels, nous n'allons retenir que trois composantes principales d'un neurone :

- Les dendrites,
- Le corps de la cellule,
- L'axone.

Les dendrites sont les branches réceptives du neurone qui transportent les signaux électriques dans le corps de la cellule. Le corps de la cellule somme et limite ces signaux entrants. L'axone est une fibre qui transmet le signal traité par le corps de la cellule vers d'autres neurones. L'information est passée entre l'axone d'un neurone à une dendrite d'un autre neurone grâce à un point de contact appelé synapse. La figure 2.6 schématise deux neurones connectés.

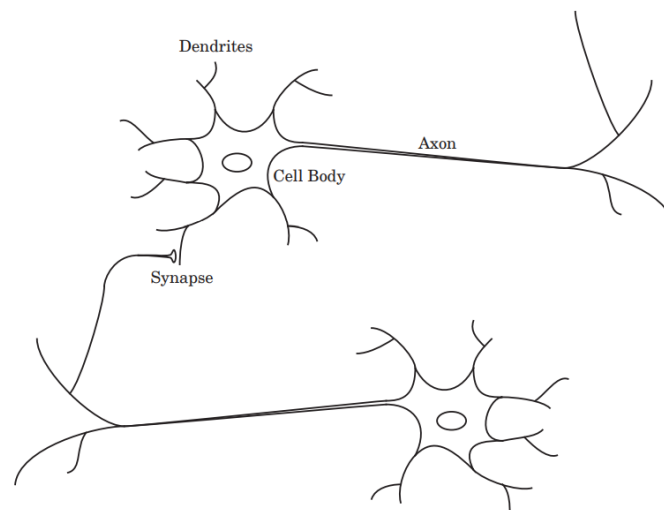


FIGURE 2.6 Schéma simplifié de deux neurones biologiques [45].

Les neurones communiquent donc entre eux via les synapses, ce qui leur permet d'échanger des informations. Ces neurones sont de ce fait connectés les uns aux autres, formant un réseau de neurones.

C'est la disposition des neurones et les forces des synapses individuelles qui déterminent la fonction du réseau de neurones. Certaines structures neuronales sont innées, et d'autres sont développées au cours de la vie en fonction du vécu. Les liaisons entre les neurones sont adaptées, modifiées, renforcées ou atténuées en fonction de nos expériences. En effet, les synapses jouent le rôle de "mémoire" entre les neurones, ce qui leur permet

d'ajuster leur fonctionnement, c'est l'apprentissage. Les structures neuronales évoluent donc en permanence.

Bien qu'un réseau de neurones artificiels (ANN pour *Artificial Neural Network*) ne soit pas aussi complexe que ceux du cerveau, cette approche permet de faire de l'apprentissage basé sur des exemples. Les réseaux de neurones artificiels sont basés sur les méthodes probabilistes et statistiques. Ils ont la faculté d'extraire des règles complexes à partir de données bruitées et imprécises. Les réseaux de neurones sont utilisés dans beaucoup de domaines. Ils permettent notamment de faire de la régression et de la classification. Par exemple, ils ont été utilisés pour faire de la classification sur des patients atteints de la maladie d'Alzheimer [46]. Autre exemple, ces réseaux ont été utilisés pour prédire la consommation de carburant des voitures pour la conduite en ville [47].

La régression a pour objectif de mettre en relation un certain nombre de variables d'entrée x avec un ensemble de résultats continus y . La classification quant à elle a pour but de classer des observations représentées par un ensemble de valeurs d'entrée x parmi des classes d'une variable cible y .

Un réseau de neurones artificiels est composé de plusieurs couches. La première couche est appelée couche d'entrée, ou *input layer* en anglais. Son rôle est de transmettre les informations d'entrées aux couches suivantes. Cette couche d'entrée est composée d'un neurone par information disponible. Ensuite, le réseau de neurones se compose de zéro à plusieurs couches qu'on appelle couches cachées, ou *hidden layers*. Pour les réseaux de neurones les plus simples (une addition par exemple), il n'y a pas besoin de couche cachée. Pour d'autres cas plus complexes, il peut y avoir un grand nombre de couches cachées. Le nombre de couches cachées et le nombre de neurones par couche cachée n'est pas déterminé par une fonction ou un théorème. Le seul moyen de trouver la meilleure solution pour un problème donné est d'essayer différentes combinaisons afin de trouver la plus optimale. Enfin, il y a la couche de sortie, ou *output layer*, dont le nombre de neurones dépend du problème, et dont le résultat est la prédiction du réseau de neurones en fonction des informations qu'il a eues en entrée. La figure 2.7 représente un schéma d'une architecture d'un réseau de neurones.

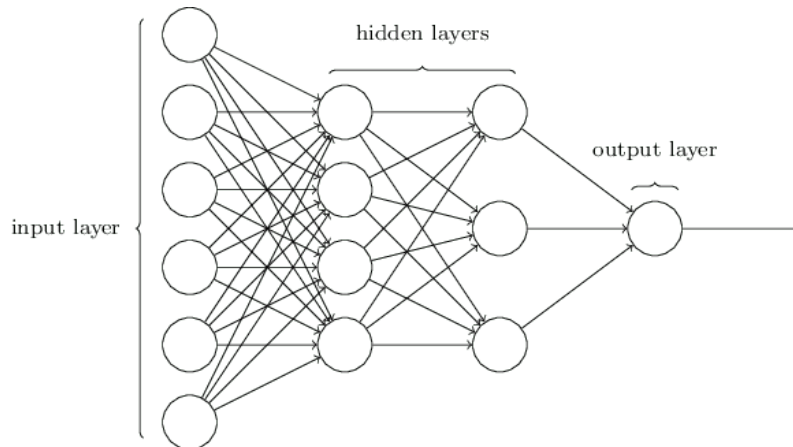


FIGURE 2.7 Architecture d'un réseau de neurones [48].

Afin d'être capable de faire des prévisions, un réseau de neurones artificiels doit être entraîné à partir de données réelles. Ces données sont constituées d'un ensemble de valeurs que l'on notera x , et du résultat réel que le réseau devrait trouver, que l'on note y . Cela permet au réseau de neurones de se corriger en cas d'erreur. C'est l'apprentissage.

Les informations transmises d'un neurone à un autre s'effectuent via une connexion avec une certaine force, c'est le poids. Ce poids permet de déterminer l'influence de l'information pour le neurone auquel elle est transmise. Plus la valeur du poids, que l'on note w , est importante, plus l'information est influente sur le neurone qui la reçoit. Chaque connexion entre deux neurones a son propre poids. Ce sont ces poids qui sont ajustés par le réseau de neurones lorsque celui-ci est en phase d'apprentissage.

Lors de la réception d'informations, le neurone fait la somme pondérée de ses entrées et y applique une fonction d'activation. La fonction d'activation détermine si le neurone va être activé ou non. La sortie s d'un neurone est représentée par la formule suivante : $f(w_1x_1 + \dots + w_nx_n)$, où f est la fonction d'activation, w est le poids de l'information et x est la valeur de l'information. La figure 2.8 schématise la structure d'un neurone.

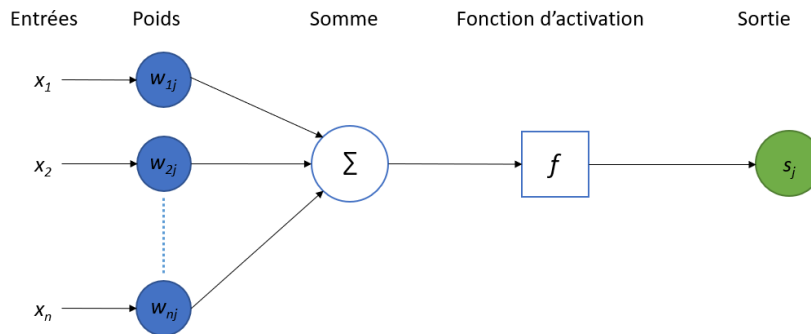


FIGURE 2.8 Structure d'un neurone artificiel.

Les réseaux de neurones les plus utilisés aujourd'hui sont les Perceptrons multicouches, dont l'architecture a été pensée par Rumelhart et McClelland [49].

Les fonctions d'activation sont les fonctions qui sont utilisées par un neurone afin de s'activer (d'envoyer l'information) à la couche de neurones suivante. Elles sont très importantes car ce sont elles qui déterminent si un neurone doit être activé ou non, en fonction de la pertinence des informations que le neurone a reçues. Il existe plusieurs fonctions d'activation, qui sont choisies en fonction de l'objectif du réseau de neurones :

- Seuil : La fonction de seuil est la fonction la plus simple. Elle renvoie 1 si la valeur x est supérieure à un seuil, et 0 sinon. Elle est utilisée dans le cas de classifieurs binaires. Quand il s'agit d'un résultat binaire, la fonction de seuil est le meilleur choix car elle va simplement activer le neurone ou non. Cette fonction est tracée sur la figure 2.9 A).
- Linéaire : Cette fonction simple est définie par $f(x) = ax$, où a est le coefficient de la fonction linéaire. La fonction linéaire est souvent utilisée pour la couche de sortie dans le cadre de la régression. Cette fonction est tracée sur la figure 2.9 B).
- Sigmoide : La fonction sigmoïde est définie par $f(x) = \frac{1}{1 + e^{-x}}$. Elle est pratique pour faire de la classification puisqu'elle permet d'obtenir une probabilité pour les différentes classes. Cette fonction est tracée sur la figure 2.9 C).
- Redresseur : Cette fonction est définie par $f(x) = \max(0, x)$. C'est l'une des fonctions les plus utilisées pour les réseaux de neurones, notamment dans les couches cachées [50]. Cette fonction est tracée sur la figure 2.9 D).
- Tangente hyperbolique : La fonction tanh est définie par $f(x) = 1 - \frac{2}{e^{2x} + 1}$. Elle est généralement meilleure que la sigmoïde. Elle convient bien aux perceptrons multicouches, et particulièrement pour les couches cachées. Cette fonction est tracée sur la figure 2.9 E).

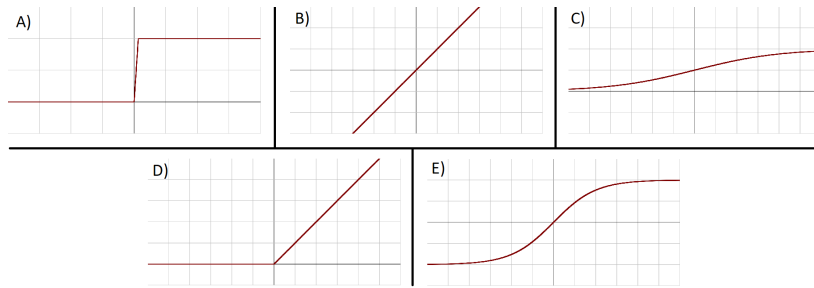


FIGURE 2.9 Fonctions d'activation. A) Seuil ; B) Linéaire ; C) Sigmoide ; D) Redresseur ; E) Tanh.

La fonction de perte, ou *loss function*, permet de comparer la prédiction en sortie \hat{y} du réseau de neurones avec le vrai résultat y qui aurait dû être prédit. Une fonction de perte souvent utilisée dans le cadre de la régression est l'erreur quadratique (perte quadratique). Pour un exemple d'apprentissage, cette fonction est définie par $(\hat{y} - y)^2$. Pour prendre en compte l'ensemble des exemples d'apprentissage, il faut calculer l'erreur quadratique moyenne (EQM), qui permet de calculer la perte quadratique moyenne pour chaque exemple. Le but est de minimiser la fonction de perte, plus la perte est petite, plus la prédiction du réseau de neurones est bonne. La formule de l'EQM est la suivante, où n représente l'ensemble d'apprentissage, Y est le vecteur des vrais résultats et \hat{Y} correspond au vecteur des résultats prédits par le réseau de neurones :

$$EQM = \frac{1}{n} \sum_{i=1}^n (\hat{Y}_i - Y_i)^2$$

La mise à jour des poids est effectuée grâce au résultat de la fonction de perte. En effet, l'algorithme du gradient permet de calculer la dérivée de la fonction de perte et d'ajuster les poids en fonction de cette dérivée. C'est la dérivée partielle de la fonction de perte qui permet de déterminer la façon dont il faut ajuster les poids de chaque neurone. L'algorithme du gradient va alors faire une rétropropagation dans le réseau de neurones afin de mettre à jour tous les poids en fonction du vecteur des dérivées partielles.

Cela permet d'améliorer les performances du modèle grâce aux exemples qui ont été fournis au réseau de neurones. Le paramétrage des poids est donc exploré au cours de l'apprentissage.

Les étapes d'entraînement d'un réseau de neurones sont les suivantes :

1. La première étape est d'initialiser tous les poids de chaque neurone avec une valeur proche de 0 (mais différente de 0).

2. Ensuite, une première observation est envoyée au réseau de neurones par la couche d'entrée, où chaque neurone reçoit une variable différente.
3. Après cela, le réseau de neurones se met en marche et la propagation "avant" (appelée *forward*) commence. Les neurones des couches cachées sont activés et la couche de sortie renvoie le résultat \hat{y} .
4. La prédiction \hat{y} est comparée avec le résultat attendu y . La rétropropagation va mettre à jour les poids en fonction du taux d'apprentissage, c'est-à-dire en fonction de l'erreur entre \hat{y} par rapport à y .
5. On réitère ensuite ces étapes en envoyant une nouvelle observation dans le réseau de neurones.

Cet entraînement peut être réalisé observation par observation (comme décrit juste avant), ou alors par lot ou mini-lot. Dans le cas de lots et de mini-lots, le réseau va mettre à jour les poids seulement après avoir traité plusieurs observations.

Afin d'optimiser le réseau de neurones, cet entraînement peut répéter plusieurs fois les étapes 2 à 5. De ce fait, l'ensemble du jeu de données est renvoyé au réseau de neurones pour qu'il ajuste ses poids. Le nombre de fois que cet entraînement est répété s'appelle le nombre d'époques.

2.2.3.2 Machine à vecteurs de support

L'objectif de l'algorithme de machine à vecteurs de support (ou SVM) [51] est de trouver un hyperplan dans un espace à N dimensions (N étant le nombre de *features*) qui classe distinctement les points de données.

Pour séparer les classes de points de données, il existe de nombreux hyperplans possibles qui pourraient être choisis. L'objectif est de trouver un plan qui a la marge maximale, c'est-à-dire la distance maximale entre les points de données des deux classes. L'optimisation de la distance de marge fournit un certain renforcement afin que les futurs points de données puissent être classés avec plus de confiance.

Les hyperplans sont des limites de décision qui aident à classer les points de données. Les points de données situés de chaque côté de l'hyperplan peuvent être attribués à différentes classes. En outre, la dimension de l'hyperplan dépend du nombre de *features*. Si le nombre de *features* en entrée est de 2, alors l'hyperplan n'est qu'une ligne. S'il est de 3, alors l'hyperplan devient un plan bidimensionnel. Il devient plus difficile d'imaginer l'hyperplan quand le nombre de *features* dépasse 3. La figure 2.10 montre un hyperplan dans un plan en 2D et un cube en 3D.

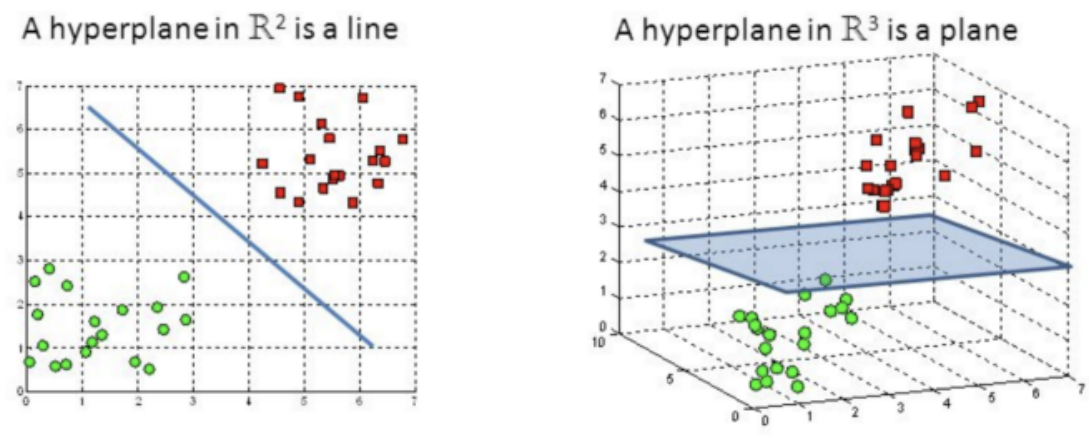


FIGURE 2.10 Hyperplans dans l'espace 2D et 3D [52].

Les vecteurs de support sont les points de données plus proches de l'hyperplan et qui influencent la position et l'orientation de l'hyperplan. En utilisant ces vecteurs de support, la marge du classifieur est maximisée. La suppression des vecteurs de support modifiera la position de l'hyperplan. Ce sont les points qui aident à construire la SVM. La figure 2.11 montre un hyperplan optimal en 2D, avec les vecteurs de support.

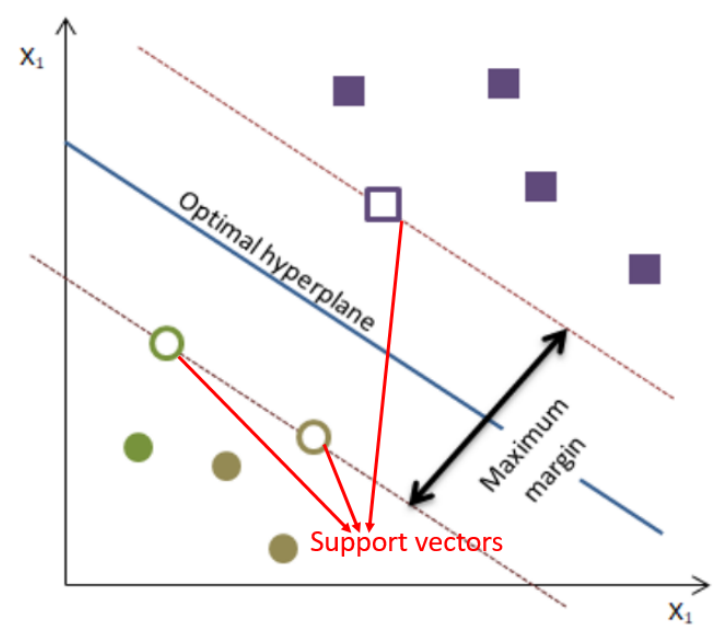


FIGURE 2.11 Hyperplan optimal [52].

Les SVM sont largement utilisées dans les tâches de classification. Par exemple, elles ont été utilisées pour faire de la classification de ravageurs (insectes) dans des serres de fraises [53]. Cependant, elles peuvent également être utilisées pour les tâches de régression en conservant les principales caractéristiques de l'algorithme (marge maximale). La régression vectorielle de support (SVR) utilise les mêmes principes que la SVM pour la classification, avec seulement quelques différences mineures.

2.2.4 Apprentissage non supervisé

L'apprentissage non supervisé est une technique d'apprentissage automatique, où le modèle n'a pas besoin d'être supervisé. En effet, le modèle apprend à partir de données non labélisées. Cela signifie que seuls les *inputs* sont disponibles, et que le modèle va donc créer ses propres *outputs* pour ces données. Le modèle va fonctionner seul dans le but de découvrir des informations dans les données qui lui sont fournies. Les algorithmes d'apprentissage non supervisé permettent d'effectuer des tâches de traitement plus complexes que l'apprentissage supervisé. Cependant, l'apprentissage non supervisé peut s'avérer être imprévisible sur les résultats obtenus.

2.2.4.1 K-Means

K-Means [54] est un algorithme de clustering. C'est une méthode de quantification vectorielle, dont l'origine vient du traitement du signal, qui vise à partitionner n observations en k clusters, dans lesquels chaque observation appartient au cluster dont le centroïde est le plus proche.

Pour cet algorithme, il est nécessaire de connaître par avance le nombre de clusters k . En effet, la première étape est de définir la valeur de k , qui est le paramètre principal de l'algorithme. Ensuite, un processus itératif est exécuté pour regrouper les données. Ce processus s'explique par les étapes suivantes :

- Etape 0 : Initialisation : k points dans l'espace sont sélectionnés aléatoirement, ils correspondent aux centroïde des clusters.
- Etape 1 : Chaque observation du jeu de données est ensuite assignée au centroïde le plus proche d'elle. Cela permet de segmenter une première fois les données en k clusters.
- Etape 2 : Le centroïde de chaque cluster est de nouveau calculé.

- Etape 3 : On répète le processus depuis l'étape 1. Le processus est répété tant que des observations sont réassignées. L'algorithme prend fin quand plus aucun changement n'est réalisable.

La figure 2.12 montre un exemple de clustering en deux clusters, dans un plan en 2 dimensions. Les exemples d'entraînement sont représentés sous la forme de points et les centroïdes des clusters sont représentés par des croix. La figure 2.12 (a) représente l'ensemble de données d'origine. La figure 2.12 (b) représente l'initialisation aléatoire des centroïdes des clusters. Lors de chaque itération, chaque exemple d'apprentissage est attribué au centre de gravité du cluster le plus proche. Ensuite le centroïde de chaque cluster est recalculé et déplacé, et le processus est répété à partir de l'étape d'assignation.

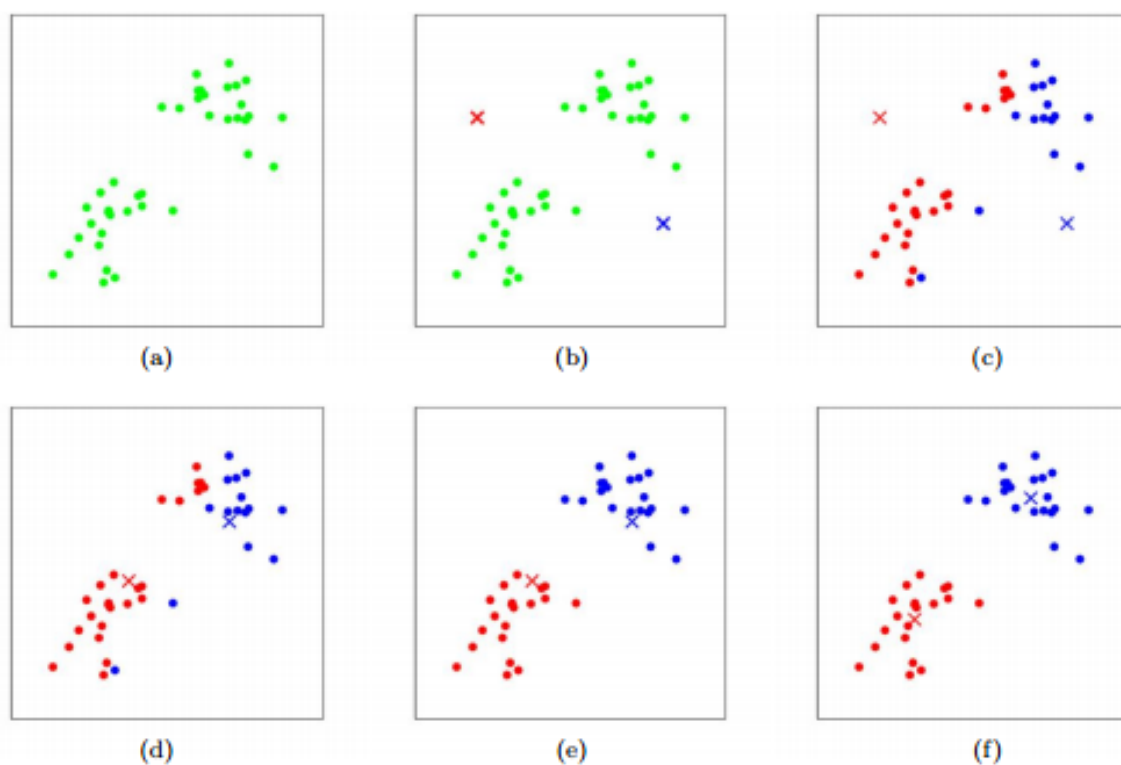


FIGURE 2.12 Illustration de clustering avec K-Means [55].

K-Means a par exemple été utilisé pour regrouper les étudiants en fonction de leur niveau d'engagement lors d'apprentissage en ligne (e-Learning) [56]

2.2.4.2 Self-Organizing Maps

Une *Self-Organizing Map* ([57], [58]), souvent appelée SOM est une technique de clustering qui permet de découvrir des catégories dans de grands ensembles de données. Les SOM sont un type de réseau de neurones artificiels qui est entraîné grâce à un apprentissage non-supervisé. L'objectif est de produire une représentation discrétisée de faible dimension (usuellement bidimensionnelle) de l'espace des *inputs* des échantillons d'apprentissage. Cette représentation est appelée carte. SOM est donc une méthode qui permet de réduire la dimensionnalité.

Les SOM diffèrent des autres réseaux de neurones artificiels car elles appliquent un apprentissage compétitif, par opposition à un apprentissage à correction d'erreurs (comme la rétropropagation avec descente de gradient). De plus, elles utilisent une fonction de voisinage pour préserver les propriétés topologiques de l'espace d'entrée.

Afin d'apprendre la forme des données, l'algorithme SOM exécute un processus itératif qui peut être résumé par les étapes suivantes :

- Etape 0 : Positionnement de manière aléatoire des neurones de la grille dans l'espace de données.
- Etape 1 : Sélection d'un point de données, soit de manière aléatoire, soit en parcourant systématiquement l'ensemble de données dans l'ordre.
- Etape 2 : Recherche du neurone le plus proche du point de données choisi. Ce neurone est appelé le *Best Matching Unit* (BMU, ou meilleure unité de correspondance en français).
- Etape 3 : Rapprocher le BMU de ce point de données. La distance parcourue par le BMU est déterminée par un taux d'apprentissage, qui diminue après chaque itération.
- Etape 4 : Rapprocher les voisins du BMU, les voisins plus éloignés se déplaçant moins. Les voisins sont identifiés à l'aide d'un rayon autour de la BMU, et la valeur de ce rayon diminue après chaque itération.
- Etape 5 : Mise à jour du taux d'apprentissage et du rayon du BMU, avant de répéter les étapes 1 à 4. Répéter ces étapes jusqu'à ce que les positions des neurones soient stabilisées.

La figure 2.13 montre comment une SOM s'adapte à un ensemble de données. La tâche bleue représente la distribution des données d'apprentissage. Le petit disque blanc est la donnée d'apprentissage actuellement tirée de cette distribution. Au début (à gauche), les nœuds SOM sont positionnés arbitrairement dans l'espace de données. Le nœud (surligné

en jaune) le plus proche de la donnée d'apprentissage est sélectionné, c'est le BMU. Il est déplacé vers la donnée d'apprentissage, comme (dans une moindre mesure) ses voisins sur la grille. Après de nombreuses itérations, la grille tend à se rapprocher de la distribution des données (à droite).

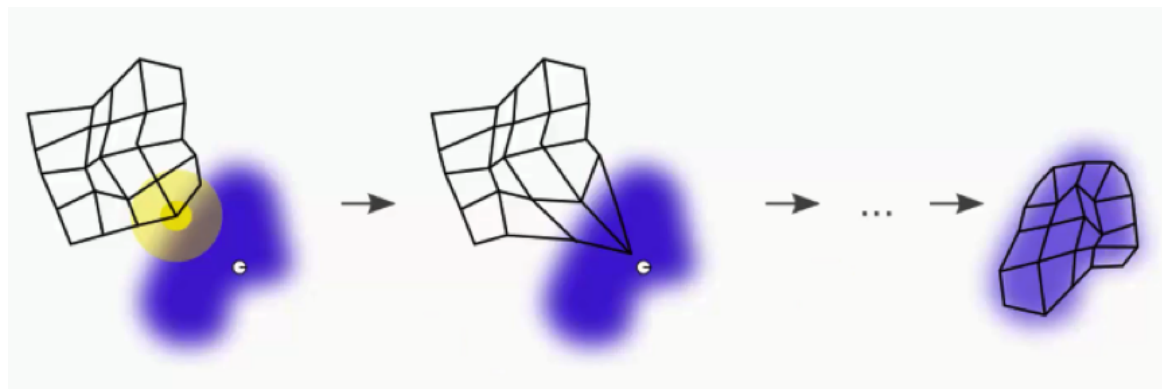


FIGURE 2.13 Illustration de la formation d'une SOM [59].

Les SOM ont par exemple été utilisées pour faire du clustering de texte [60].

2.2.4.3 Density-Based Spatial Clustering of Applications with Noise

Density-based spatial clustering of applications with noise ([61], [62]), plus couramment appelé DBSCAN est un algorithme de clustering de données qui est utilisé dans l'exploration de données et l'apprentissage automatique, spécialement en apprentissage non-supervisé.

En utilisant un ensemble de points, DBSCAN regroupe les points proches les uns des autres en fonction d'une mesure de distance (généralement une distance euclidienne) et d'un nombre minimum de points. Il détecte également comme valeurs aberrantes les points situés dans des régions à faible densité.

Il y a deux paramètres principaux dans l'algorithme DBSCAN :

- *eps* : Spécifie à quelle distance les points doivent être proches les uns des autres pour être considérés comme faisant partie d'un cluster. Cela signifie que si la distance entre deux points est inférieure ou égale à cette valeur (*eps*), ces points sont considérés comme voisins.
- *minPts* : Correspond au nombre minimum de points pour former une région dense. Par exemple, si nous définissons ce paramètre avec la valeur 5, nous avons besoin d'au moins 5 points pour former une région dense.

Sur la base de ces deux paramètres, les points sont classés en tant que point central (*core point*), point frontière (*border point*) ou valeur aberrante (*outlier*) :

- Point central : Un point est un point central s'il y a au moins $minPts$ points (y compris le point lui-même) dans sa zone environnante avec un rayon eps .
- Point frontière : Un point est un point frontière s'il est accessible à partir d'un point central et qu'il y a moins de $minPts$ points dans sa zone environnante.
- Valeur aberrante : un point est une valeur aberrante s'il ne s'agit pas d'un point central et qu'il n'est accessible à partir d'aucun point central.

Ces différents points peuvent être visualisés sur la figure 2.14. Dans ce cas, $minPts$ vaut 4. Les points rouges sont des points centraux car il y a au moins 4 points dans leur zone environnante avec un rayon eps . Cette zone est représentée par les cercles sur la figure. Les points jaunes sont des points frontières car ils sont accessibles à partir d'un point central et ont moins de 4 points dans leur voisinage. Un point accessible signifie qu'il se situe dans la zone environnante d'un point central. Les points B et C ont deux points (y compris le point lui-même) dans leur voisinage (c'est-à-dire la zone environnante avec un rayon eps). Enfin N est une valeur aberrante car ce n'est pas un point central et il ne peut pas être atteint à partir d'un point central.

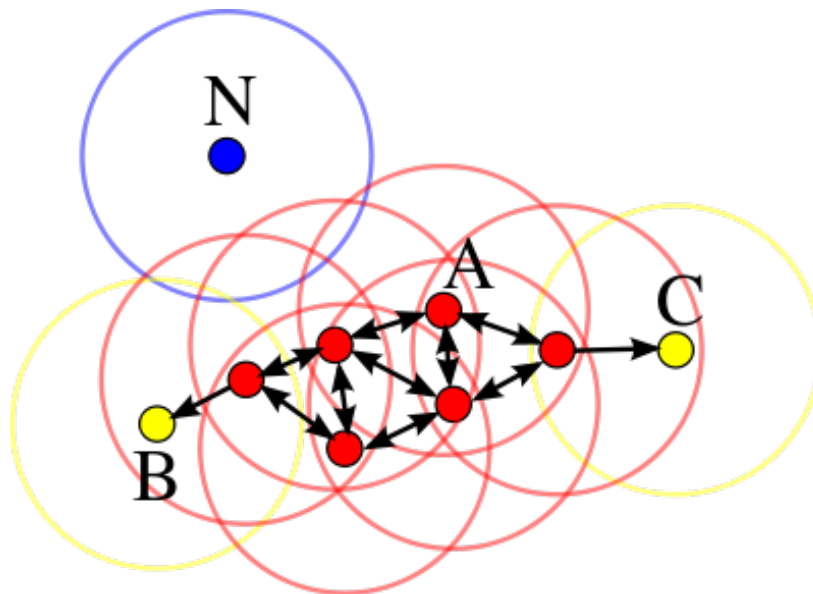


FIGURE 2.14 Schéma des différents points de l'algorithme DBSCAN avec $minPts = 4$ [63].

Au début de l'algorithme, un point de départ est sélectionné au hasard et sa zone de voisinage est déterminée à l'aide du rayon eps . S'il y a au moins $minPts$ points dans le voisinage, le point est marqué comme point central et la formation d'un cluster commence.

Sinon, le point est marqué comme du bruit. Une fois qu'une formation de cluster commence (disons le cluster *A*), tous les points dans le voisinage du point initial deviennent une partie du cluster *A*. Si ces nouveaux points sont également des points centraux, les points qui se trouvent à proximité sont également ajoutés au groupe *A*. Bien entendu, un point, initialement marqué comme du bruit, peut être retraité par l'algorithme après qu'un cluster ait commencé à se former, et faire partie d'un nouveau cluster.

L'étape suivante consiste à choisir au hasard un autre point parmi les points qui n'ont pas été traités lors des étapes précédentes. Ensuite, la même procédure s'applique. Ce processus est terminé lorsque tous les points sont traités.

En appliquant ces étapes, l'algorithme DBSCAN est capable de trouver des régions à haute densité et de les séparer des régions à faible densité. Un cluster est constitué des points centraux voisins (c'est-à-dire accessibles les uns des autres) et de tous les points frontières de ces points centraux. La condition requise pour former un cluster est d'avoir au moins un point central.

DBSCAN a par exemple été utilisé pour réaliser un clustering de galaxies [64].

2.2.4.4 Hierarchical Clustering

Le clustering hiérarchique [65] (*hierarchical clustering* en anglais, aussi connu sous le nom de *hierarchical cluster analysis*) est un algorithme qui regroupe des éléments similaires en *clusters*. Le point de terminaison est un ensemble de *clusters*, où chaque *cluster* est distinct des autres, et les éléments au sein de chaque *cluster* sont globalement similaires les uns aux autres.

Le clustering hiérarchique peut être utilisé avec une matrice de distance ou des données brutes. Lorsque des données brutes sont fournies, l'algorithme calcule automatiquement une matrice de distance en arrière-plan.

Le clustering hiérarchique commence par traiter chaque observation comme des *clusters* distincts. Ensuite, il exécute à plusieurs reprises les deux étapes suivantes :

1. Identifier les deux *clusters* les plus proches,
2. Fusionner les deux *clusters* les plus similaires.

Cela continue jusqu'à ce que tous les *clusters* soient fusionnés. Ceci est illustré dans les schémas de la figure 2.15, qui montre comment l'algorithme fusionne des clusters ensemble.

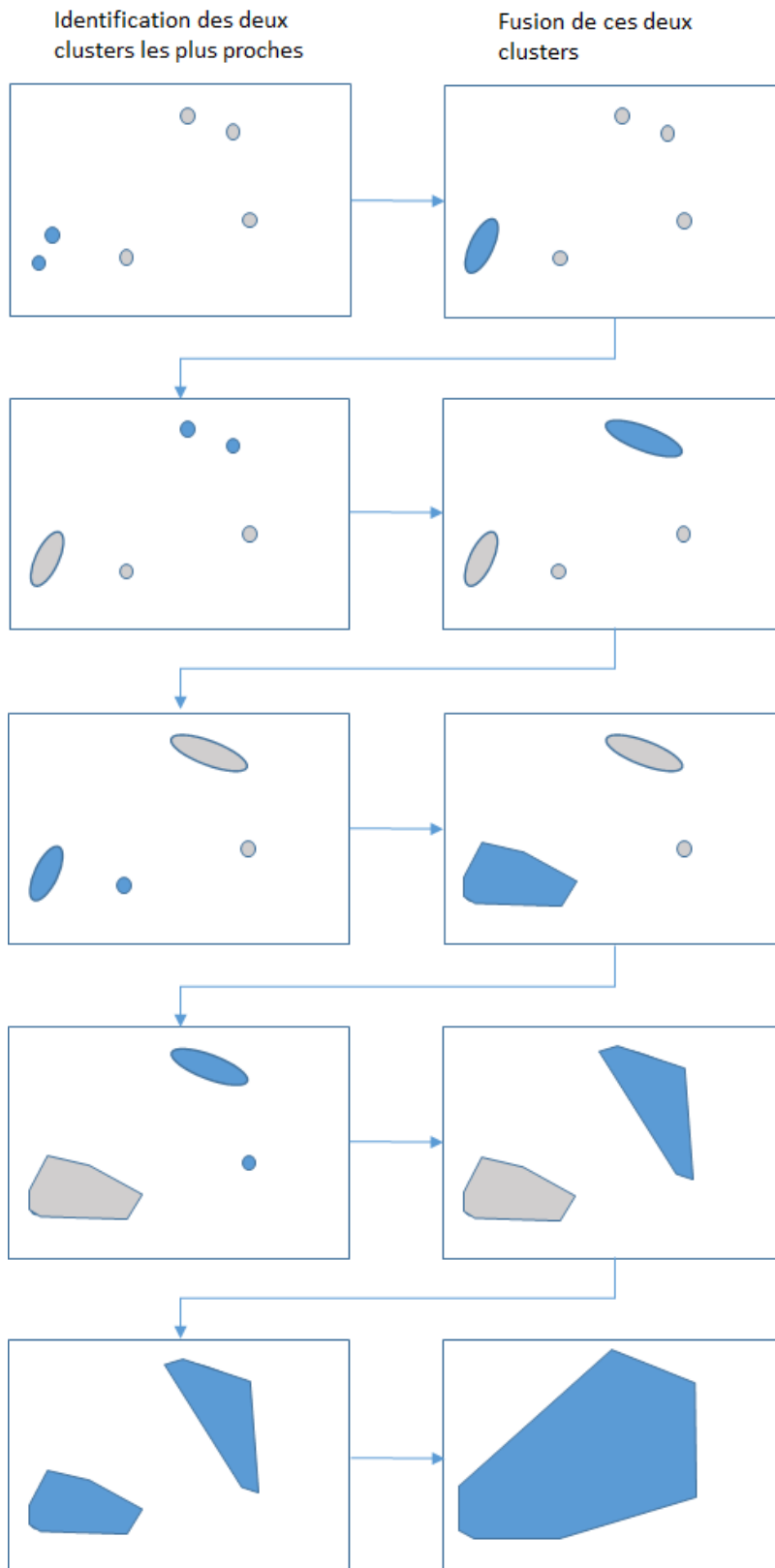


FIGURE 2.15 Schéma du processus de clustering hiérarchique.

La sortie principale du clustering hiérarchique est un dendrogramme, qui montre la relation hiérarchique entre les clusters, comme le présente la figure 2.16. Un dendrogramme est un diagramme qui montre la relation hiérarchique entre les objets. L'utilisation principale d'un dendrogramme est de trouver la meilleure façon de regrouper des éléments aux clusters. La clé de l'interprétation d'un dendrogramme est de se concentrer sur la hauteur à laquelle deux objets sont réunis. Dans l'exemple de la figure 2.16, nous pouvons voir que A et B sont les plus similaires, car la hauteur du lien qui les relie est la plus petite. Les deux éléments suivants les plus similaires sont E et F.

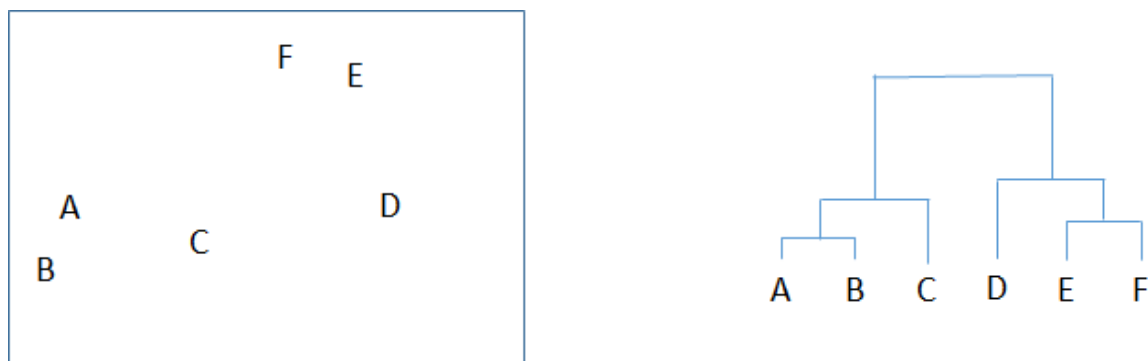


FIGURE 2.16 Exemple de Dendrogramme.

Un critère important est la mesure de distance (similitude) entre les éléments. Dans l'exemple ci-dessus, la distance entre deux clusters a été calculée en fonction de la distance Euclidienne. Cependant, il existe d'autres mesures de distance. Le choix de la métrique de distance doit être fait sur la base des préoccupations théoriques du domaine d'étude. Autrement dit, une métrique de distance doit définir la similitude d'une manière qui est sensible pour le domaine d'études.

Après avoir sélectionné une métrique de distance, il est nécessaire de déterminer le critère de liaison, c'est-à-dire d'où la distance est calculée. De nombreux critères de liaison ont été développés, mais les plus communs sont les suivants :

- *Single* utilise le minimum des distances entre toutes les observations des deux ensembles,
- *Complete* ou *Maximum linkage* utilise les distances maximales entre toutes les observations des deux ensembles,
- *Average* utilise la moyenne des distances de chaque observation des deux ensembles,
- *Ward* minimise la variance des clusters fusionnés.

Le clustering hiérarchique fonctionne généralement en fusionnant séquentiellement des clusters similaires. C'est ce que l'on appelle *agglomerative hierarchical clustering*. En théorie, cela peut également être fait en regroupant initialement toutes les observations dans un cluster, puis en divisant successivement ces clusters. C'est ce que l'on appelle le *divisive hierarchical clustering*. Le clustering par division est rarement effectué en pratique.

Les observations sont attribuées aux clusters en traçant une ligne horizontale à travers le dendrogramme. Les observations réunies sous la ligne sont regroupées dans le même cluster. La figure 2.17 montre un exemple de dendrogramme.



FIGURE 2.17 Exemple de dendrogramme avec ligne de séparation des clusters.

Dans cet exemple il y a donc deux clusters, mais ceci est sujet à l'interprétation faite par l'expert et donc peut varier en fonction des interprétations.

Par exemple, Hierarchical Clustering a été utilisé pour regrouper les différentes démarches humaines, analysées grâce à des électromyogrammes [66].

2.2.5 Classification et clustering

La classification et le clustering sont deux méthodes d'identification de *patterns* utilisées dans l'apprentissage automatique. Bien que les deux techniques présentent certaines similitudes, il existe une différence majeure entre elles. En effet, la classification utilise des classes prédéfinies dans lesquelles les objets sont labélisés, tandis que le clustering identifie les similitudes entre les objets, qu'il regroupe en fonction de caractéristiques communes et qui les différencie des autres groupes d'objets. Ces groupes sont appelés *clusters*.

Dans le domaine du *Machine Learning*, le clustering fait partie de l'apprentissage non supervisé, ce qui implique que le résultat attendu est inconnu. La classification quant à elle

appartient à l'apprentissage supervisé, ce qui signifie que la sortie possible de l'algorithme est connue et permet d'évaluer le modèle.

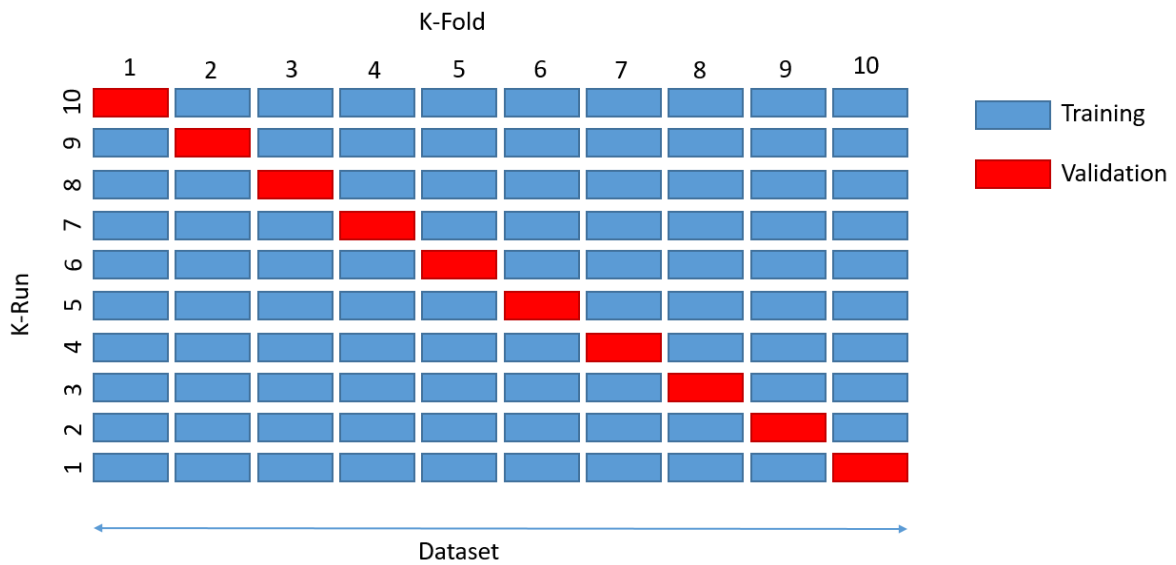
2.2.6 K-fold cross validation

En *Machine Learning*, le jeu de données est divisé en deux : un ensemble d'entraînement et un ensemble de test, aussi appelé de validation [67]. L'ensemble de données d'entraînements est utilisé, comme son nom l'indique, pour entraîner le modèle. Cet ensemble représente en général entre 75% et 80% du jeu de données global. L'ensemble de validation quant à lui est utilisé pour évaluer le modèle après la phase d'entraînement. Cet ensemble contient le reste des données, soit en général entre 20% et 25% du jeu de données global. Il est important de séparer les données en deux ensembles, car cela permet d'évaluer le modèle sur des données qui n'ont pas été utilisées pour l'entraîner, ce qui permet donc d'avoir une évaluation qui n'est pas biaisée [68].

Cependant, ce n'est pas nécessairement la technique la plus optimale. L'une des raisons est la variance qu'il peut exister lors de l'entraînement du modèle. En effet, entraîner deux modèles identiques avec le même ensemble de données ne donnera pas forcément le même résultat. De ce fait, il est nécessaire de savoir si l'entraînement du modèle est fiable, donc avec une variance faible ou au contraire si la variance est élevée auquel cas le modèle n'est pas bon.

Par conséquent, la validation croisée est utilisée pour diviser les données en k sous-ensembles [69]. *K-Fold Cross Validation* est un algorithme itératif, qui itère k fois. A chaque itération $k - 1$ sous-ensembles vont être utilisés pour entraîner le modèle et le dernier sous-ensemble est utilisé pour évaluer les performances du modèle. Le sous-ensemble permettant d'évaluer le modèle est différent à chaque itération [70]. Ceci permet donc de connaître la variance du modèle et de savoir s'il est performant.

La figure 2.18 schématise *K-Fold Cross Validation* avec $k = 10$. Chaque ligne représente une itération et chaque colonne représente un sous-ensemble de données. Pour chaque itération, les sous-ensemble bleus sont utilisés pour l'entraînement du modèle et le sous-ensemble rouge est utilisé pour évaluer le modèle.

FIGURE 2.18 Schéma de K -Fold Cross Validation avec $k = 10$.

2.2.7 Bilan

Nous avons vu dans cette partie que l'exploration de données permet d'extraire de l'information d'une masse de données. Nous avons vu qu'il existe deux grandes approches permettant de faire de l'apprentissage automatique :

- L'apprentissage supervisé,
- L'apprentissage non-supervisé.

Le choix de l'une ou de l'autre approche dépend directement de la connaissance déjà disponible sur l'ensemble de données d'apprentissage. Nous avons également pu voir que pour chaque approche, plusieurs algorithmes existent, chacun avec ses avantages et ses inconvénients.

Chapitre 3

Expérimentation

Ce chapitre a pour objectif de présenter l'expérimentation que nous avons mise en place sur la digue de l'Artha afin d'étudier la puissance d'impact des vagues.

Pour cela, nous allons commencer par décrire la digue de l'Artha et l'histoire de la baie de Saint-Jean-de-Luz.

Ensuite, nous présenterons la station que nous avons installée sur la digue de l'Artha, en détaillant toutes les parties de celle-ci : capteurs, station de mesure, stockage, alimentation, algorithmes, format des données, etc.

Enfin, nous finirons par présenter les données qui proviennent d'autres stations, que nous avons utilisées pour nos travaux.

3.1 Besoins

Etudier un phénomène, quel qu'il soit, nécessite d'avoir un volume de données suffisamment large. En effet, un gros volume de données permet d'étudier le phénomène dans sa globalité. Cependant, peu de données sont disponibles sur la pression d'impact des vagues et encore moins en période de tempête, lorsque les conditions sont extrêmes. La raison principale de ce manque de données est l'environnement hostile dans lequel les structures côtières sont très souvent situées. De ce fait, placer une station de mesure sur une structure côtière s'avère complexe et coûteux.

Certains l'ont fait, comme De Rouville en 1938 [29]. Cependant, à cause de la technologie de cette époque, la collecte de mesures avait été effectuée sur une période relativement courte.

D'autres ont étudié le phénomène en laboratoire dans des bassins à houles ([37], [32], [71], [36], [72]). Cependant, les études en laboratoire ne prennent pas en compte toutes les variables environnementales, comme le courant ou le vent, qui influent sur le phénomène et donc peuvent le modifier. De plus, il y a une problématique de mise à l'échelle entre les études en laboratoire et le phénomène réel.

C'est pour ces raisons que nous avons choisi de mettre en place une station de mesure sur une structure côtière. L'objectif est d'avoir une station qui fonctionne en permanence dans le but de pouvoir récolter des données de pressions d'impact dans toutes les conditions météorologiques possibles, que ce soit en période de tempête ou en période calme. Grâce à cela, nous pouvons acquérir des données en continu afin d'avoir un gros échantillon de données variées sur lequel travailler. Cela a pour but de permettre aux océanographes d'étudier plus précisément les dynamiques des vagues et les pressions d'impact des vagues sur les ouvrages côtiers. De plus, cela permet aussi de pouvoir suivre l'évolution du phénomène et donc à terme de faire de la supervision de la structure côtière concernée.

3.2 Station de l'Artha

La station expérimentale a été installée sur la digue de l'Artha. L'objectif est de mesurer les pressions d'impact des vagues sur cet ouvrage côtier. Ces mesures permettent de récolter, traiter et analyser un grand nombre de données de pressions d'impact de vague sur la digue. La première version de la station est présentée dans [73], [74] et [75]. Cette première version était équipée de seulement deux capteurs de pression, alignés sur l'axe vertical permettant d'enregistrer les pressions d'impact de vagues en 1D. Dans la suite de ce chapitre nous allons présenter la nouvelle station que nous avons mise en place, composée de 16 capteurs de pression disposés sur l'axe vertical et sur l'axe horizontal, permettant d'enregistrer les pressions d'impact en 2D.

3.2.1 Histoire de la baie de Saint-Jean-de-Luz

Le Pays Basque a de tout temps été frappé par de fortes tempêtes. Les villes situées sur le littoral en ont beaucoup souffert, comme la ville de Saint-Jean-de-Luz qui a fréquemment été touchée par des houles destructrices. Les luziens ont connu beaucoup de raz-de-marée à cause de l'effondrement de la protection naturelle de la baie au XVIIIème siècle [76].

Il aura fallu attendre 1783 pour que Louis XVI établisse un plan de sauvegarde impliquant la fermeture de la rade. Le but de ce projet est de construire deux digues, l'une depuis Socoa et l'autre depuis Sainte-Barbe. Les deux digues seront séparées par un goulet, permettant aux navires de passer. C'est donc cinq ans plus tard, en 1788, que la construction de ces deux digues prend fin. Elles mesurent chacune 175 mètres de long pour 9 mètres de hauteur. Ces structures ne vont cesser d'être détruites par la houle et reconstruites, jusqu'en 1852 où une tempête dévasta un quart de Saint-Jean-de-Luz.

En 1854, sensible à la situation des habitants de Saint-Jean-de-Luz, Napoléon III mis en place un programme de construction de trois structures côtières permettant la protection de la ville : la digue de Socoa, la digue de Sainte-Barbe et la digue de l'Artha. Cette dernière sera construite entre les deux autres, sur un haut-fond. La digue de Socoa fût terminée en 1876, celle de Sainte-Barbe en 1888 et enfin celle de l'Artha ne fût terminée qu'en 1895 dû à son emplacement en plein milieu de la baie. Ces trois structures, bien qu'appelées digues, sont en réalité des brises-lames comme nous avons pu le voir dans la section 1.2. En effet, en français le mot digue est souvent utilisé par abus de langage pour désigner un autre type de structure côtière.

La digue de l'Artha se trouve à 250 mètres de la digue de Socoa et à 390 mètres de la digue de Sainte-Barbe et elle mesure 250 m (mètres linéaires). La figure 3.1 est une photo sur laquelle les trois digues sont visibles.



FIGURE 3.1 Photo de la baie de Saint-Jean-de-Luz où les trois digues sont présentes. La première digue est la digue de Sainte-Barbe, celle au milieu de l'eau est la digue de l'Artha, et enfin celle du fond est la digue de Socoa.

Malgré cela, la ville de Saint-Jean-de-Luz connut plusieurs tempêtes, dont la plus importante eut lieu en 1951. Ci-dessous, nous pouvons voir quelques photos des dégâts causés par cette tempête [77].



FIGURE 3.2 Photos de 1951 montrant les dégâts de la tempête à Saint-Jean-de-Luz [77].



FIGURE 3.3 Photo de 1951 montrant un bloc de 40 tonnes sur la digue [77].

Ce genre de grosses tempêtes et de grosses vagues se produisent toujours. En effet, encore récemment une vague a déposé un bloc de 50 tonnes sur la digue de l'Artha [78]. Les figures 3.4 et 3.5 montrent des photos de ce bloc.



FIGURE 3.4 Photo du 28 février 2017 montrant une vague déposant un bloc de 50 tonnes sur la digue de l'Artha [78].



FIGURE 3.5 Photo de mars 2017 montrant le bloc de 50 tonnes sur la digue de l'Artha [78].

3.2.2 La station

Dans cette section nous allons décrire les différents éléments qui constituent la nouvelle station que nous avons mise en place sur la digue de l'Artha. Cette amélioration est nécessaire car elle permet d'enregistrer les pressions d'impact sur plus de points de la structure et donc de mieux visualiser la manière dont les vagues percutent la digue.

3.2.2.1 Les capteurs

Nous avons équipé la station de 16 capteurs de pressions piézorésistifs. Un capteur piézorésistif est un capteur qui est plongé dans un liquide permettant de le protéger. Une membrane en acier inoxydable sépare ce liquide du fluide dont la pression est à mesurer. C'est la déformation de cette membrane sous l'effet de la pression extérieure qui engendre une variation de la pression du liquide qui se trouve autour du capteur. Le capteur délivre un signal proportionnel à cette pression, converti en signal de sortie. La figure 3.6 schématise un tel capteur.

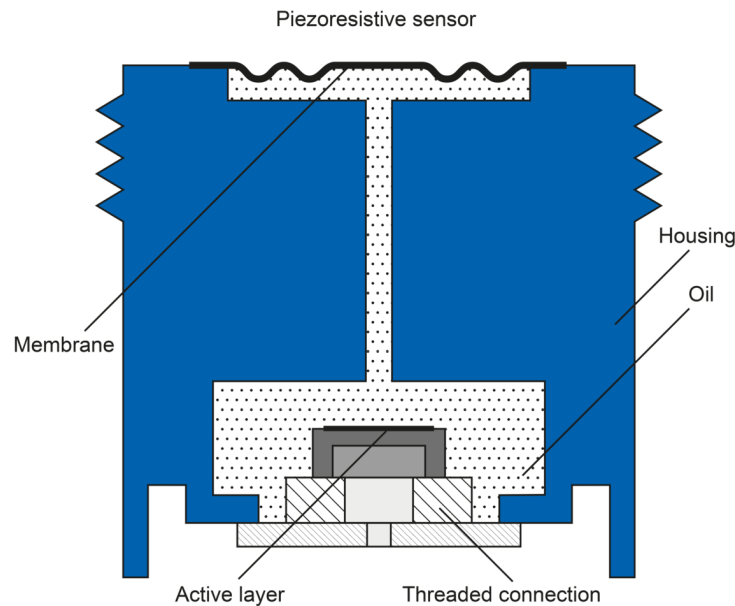


FIGURE 3.6 Schéma d'un capteur piézorésistif [79].

Les capteurs que nous avons placés sont des transmetteurs de pression absolue piézorésistifs (référence Keller© PAA-25). La figure 3.7 est un tableau qui donne les caractéristiques principales des capteurs. La fiche technique complète est disponible en annexe D.

Specification	Value
Signal output/Type	4...20mA /2-Wire
Frequency response	1 kHz
Range (FS=full scale)	0-5 bar
Typical error	±0.2% FS
Protection	IP68

FIGURE 3.7 Caractéristiques principales du capteur utilisé.

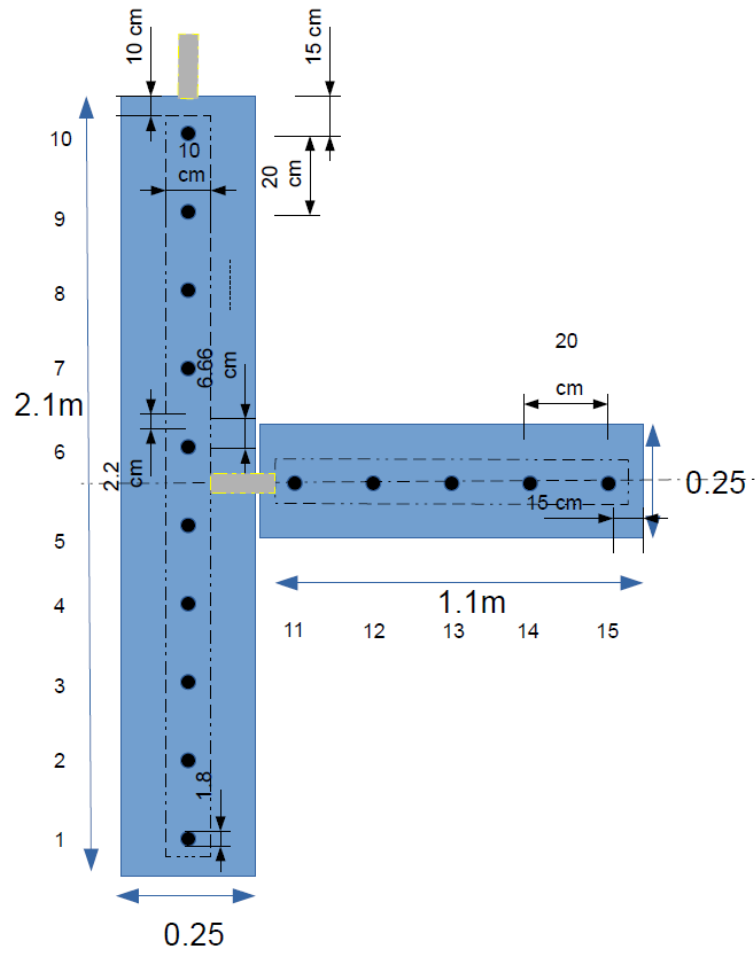
Les capteurs présentent une fonction de transfert linéaire entre la sortie de courant électrique (4..20mA) et la pression mesurée en bar. Cette fonction est la suivante :

$$P = P_M(I - K)/(K1 - K)$$

où $P_M = 5$ bars (étendue de mesures de la pression en bar du capteur), $K = 4.10^{-3}$ A et $K1 = 20.10^{-3}$ A (sorties du signal).

L'objectif était de placer les capteurs à l'endroit où les plus grosses vagues sont censées percuter la digue. Cette position a été déterminée grâce à l'observation visuelle des dommages récurrents combinée à l'analyse statistique des vagues [74]. De ce fait, les capteurs ont été positionnés sur le musoir ouest de la digue.

Initialement, seulement deux capteurs avaient été installés. Ces capteurs étaient alignés verticalement, séparés par une distance de 1.90m. Ceci permettait d'enregistrer les pressions d'impact en 1D. Nous avons fait évoluer cette installation afin d'avoir plus de capteurs. La station compte maintenant 16 capteurs, positionnés en forme de croix permettant désormais d'enregistrer les phénomènes d'impact en 2D. La figure 3.8 est un schéma qui représente le positionnement des capteurs. Chaque capteur est nommé grâce à un identifiant. Le capteur 0 correspond à celui du bas, puis cet identifiant est incrémenté en montant sur l'axe vertical, pour arriver au capteur 10 qui est situé le plus haut. Ensuite, il y a 6 capteurs sur un axe horizontal, qui sont numérotés de 11 à 15. Il faut noter que seuls 15 capteurs sont présents sur la croix, le dernier étant le capteur 0 qui se situe plus bas, légèrement sur la gauche. Ce capteur 0 correspond au capteur du bas de la première version de la station. Les figures 3.9 et 3.10 sont des photos de ces capteurs sur le musoir ouest de la digue. Sur la photo 3.9 nous pouvons bien voir le capteur 0, entouré en rouge.



0

FIGURE 3.8 Schéma du positionnement des capteurs de pression.



FIGURE 3.9 Photo rapprochée du mur avec les capteurs. Le capteur 0 est entouré en rouge.



FIGURE 3.10 Photo panoramique du musoir ouest avec les capteurs.

Ces capteurs sont reliés à une station de mesure qui se trouve 50m plus loin, derrière la digue à l'abri des vagues.

3.2.2.2 Station de mesure

Une station de mesure est un équipement permettant de mesurer et collecter des données provenant de capteurs. Ici, la station de mesure est connectée aux 16 capteurs de pression.

Nous utilisons un CompactRIO (ou Crio) de National Instruments© [80]. Le CompactRIO est la combinaison d'un contrôleur en temps réel, de modules I/O (*Input/Output*, ou Entrée/Sortie en français) reconfigurables (RIO), d'un module FPGA (*Field Programmable*

Gate Array) et d'un châssis d'extension Ethernet. Les contrôleurs temps réel CompactRIO comprennent un microprocesseur pour la mise en œuvre d'algorithmes de contrôle et prennent en charge une large gamme de fréquences d'horloge. Les contrôleurs ne sont compatibles qu'avec les modules d'E/S de la série C. Le module FPGA peut être utilisé pour implémenter un traitement de données haute performance sur un système reconfigurable. Le FPGA peut être programmé séparément et est connecté au contrôleur en temps réel à l'aide d'un bus PCI interne. Le châssis Ethernet comprend des ports Ethernet permettant de connecter le contrôleur CompactRIO à un PC ou à d'autres équipements. Les contrôleurs CompactRIO peuvent être programmés avec LabVIEW (le langage de programmation graphique de National Instruments®), C, C++, ou Java. LabVIEW doit obligatoirement être utilisé pour programmer le FPGA intégré. Le CompactRIO fonctionne sur un système d'exploitation temps réel basé sur Linux, NI Linux Real-Time. Les programmes créés dans LabVIEW sont compilés en code machine pour NI Linux Real-Time.

En raison du caractère isolé de la digue de l'Artha, il était important de choisir un système d'acquisition de données flexible pouvant être programmable à distance. C'est pour cela qu'un CompactRIO 9076 de National Instruments® a été utilisé. Ce modèle a également pour avantage d'être robuste, ce qui est nécessaire pour notre besoin. Ce système d'acquisition peut contrôler 4 modules (NI-9203 \pm 20 mA, 8 voies, 16 bits) où 8 capteurs peuvent être branchés sur chaque module. Cela signifie que nous pourrions par exemple mettre en place jusqu'à 32 capteurs de pression de type Keller® PAA25.

3.2.2.3 Stockage

Le CompactRIO que nous utilisons est équipé d'un port USB. Cela nous permet de brancher une clé USB de 128 Go sur laquelle les données sont enregistrées. Cette clé USB permet de faire tampon entre le moment où les données sont enregistrées par la station de mesure et le moment où nous les rapatrions sur nos serveurs.

3.2.2.4 Système de transmission

Pour plusieurs raisons, nous avons besoin d'avoir accès à la station à distance. Dans ce but, nous avons équipé la station d'un routeur 4G, qui est connecté au CompactRIO par Ethernet. Un réseau privé virtuel (VPN) est configuré pour assurer un canal de communication fiable avec le système d'acquisition de données.

Cela nous permet de réaliser plusieurs choses à distance :

- Vérifier le bon fonctionnement de la station,
- Modifier à distance le paramétrage de la station,

- Changer le programme embarqué,
- Consulter le nombre d'enregistrements actuellement stockés sur la station,
- Rapatrier les enregistrements.

La seule faiblesse de ce système de transmission 4G est la perte de signal en période de tempête. En effet, nous utilisons la 4G qui est un réseau cellulaire. Ce type de réseaux peut être perturbé lors d'intempéries, ce qui entraîne une perte de l'accès à la station à distance pendant une courte période. En revanche, les données sont malgré tout enregistrées sur la clé USB, évitant qu'elles soient perdues. Ces données sont ensuite rapatriées lorsque le signal est retrouvé.

3.2.2.5 Alimentation

La digue de l'Artha est un site isolé sur lequel aucune arrivée électrique n'est possible. Par conséquent, il était nécessaire de concevoir une installation pour fournir de l'énergie électrique aux appareils de mesure et au système de communication de données. Pour cela, un système photovoltaïque est installé sur la digue, à l'abri des vagues. L'alimentation se compose de trois composants :

- Un panneau photovoltaïque (SolarWorld© 150Wc),
- Un régulateur de charge solaire (Steca Solarix© PRS 1010 12 / 24V 10A),
- Une batterie haute capacité (Banner© SBV 200 -12V 200Ah).

Sur la figure 3.11 nous pouvons voir dans le cercle rouge la plaque qui protège la station ainsi que le panneau photovoltaïque.



FIGURE 3.11 Photo de la digue de l'Artha où se situe la station et le panneau photovoltaïque.
N.B. : Les capteurs sont sur l'autre pan de mur du musoir peint en blanc.

3.2.2.6 Algorithme d'acquisition

L'algorithme d'acquisition est l'algorithme qui permet de lancer des acquisitions de mesures de pressions sur la digue de l'Artha, d'enregistrer les données dans un fichier et d'envoyer ce fichier sur le serveur de l'université. La figure 3.12 schématise cet algorithme.

Pour chaque acquisition il est possible de modifier les paramètres de mesures grâce à des Web Services. Les principaux paramètres sont la fréquence d'échantillonnage Fe en Hz et la durée de l'acquisition $duration$ en minutes. Par défaut, ces paramètres sont fixés avec $Fe = 10kHz$ et $duration = 10min$. Trois modes d'acquisition sont possibles : automatique, continu et manuel. En mode automatique, la station va lancer une acquisition à chaque début d'heure ($t_{start}=XX :00 :00$), à une fréquence d'échantillonnage Fe , pour une durée $duration$. En mode continu, la station va enregistrer en permanence à une fréquence d'échantillonnage Fe jusqu'à ce que l'acquisition soit stoppée manuellement. Enfin en mode manuel, l'acquisition est lancée manuellement au moment t_{start} choisi par l'utilisateur, avec une fréquence d'échantillonnage Fe pour une durée $duration$.

Quel que soit le mode, au départ un fichier de mesures est créé. C'est dans ce fichier que toutes les mesures vont être écrites. Ensuite, le FPGA est mis en route, permettant l'acquisition des données. La variable $nb_samp_sup_thres$ est initialisée avec la valeur 0. Cette variable permet de compter le nombre d'échantillons supérieurs à un certain seuil, comme expliqué plus loin dans la section 3.2.2.10.

L'enregistrement commence, et les données récoltées sont écrites dans le fichier de mesures. A chaque pas de temps t_{mes} , l'échantillon du capteur 0 ($c0$) est comparé avec le seuil. Si la valeur de cet échantillon est supérieure au seuil, alors la variable $nb_samp_sup_thres$ est incrémentée de 1. Ensuite, l'algorithme vérifie si l'acquisition est terminée ou s'il faut encore enregistrer des données. Pour cela, t_{mes} est comparé à t_{expe} , qui correspond au temps où l'acquisition doit s'arrêter ($t_{expe} = t_{start} + duration$). Si t_{mes} n'est pas égale à t_{expe} , alors l'algorithme boucle et enregistre un nouvel échantillon, sinon l'acquisition prend fin et le FPGA est arrêté.

Ensuite, le fichier de mesures est traité. Si $nb_samp_sup_thres$ n'est pas supérieur à 0, cela signifie qu'il n'y a rien d'intéressant dans cet enregistrement et il est supprimé. Si au contraire, cette variable est supérieure à 0, alors le fichier de mesures est converti en binaire puis compressé en ZIP afin de réduire le volume du fichier. Le fichier ZIP est ensuite envoyé via FTP au serveur de l'université. En cas de réussite de transfert, le fichier est supprimé de la station. En cas d'échec du transfert, le fichier est stocké en local, en attendant un transfert manuel.

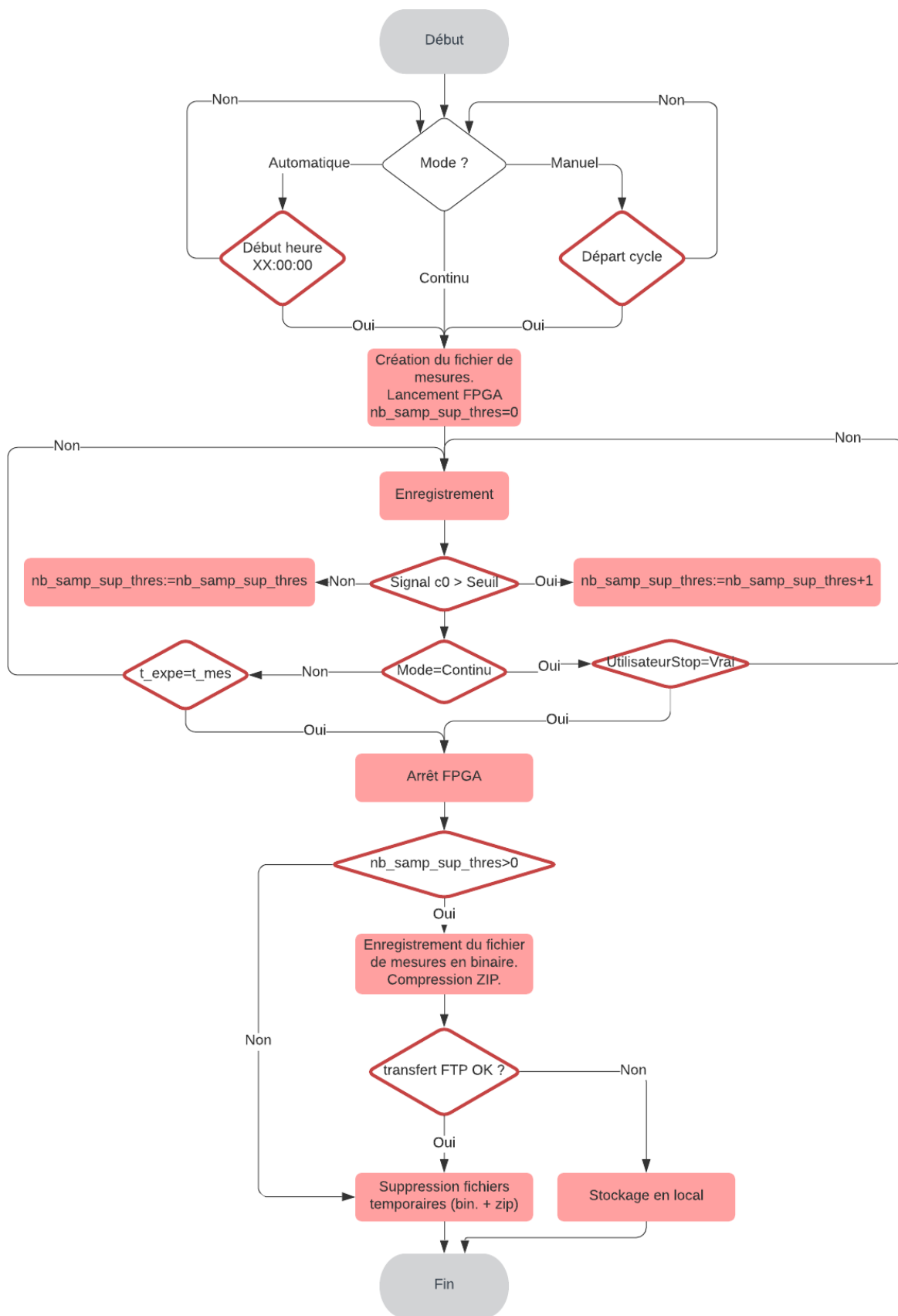


FIGURE 3.12 Algorithme de l'algorithme d'acquisition des mesures de pression.

3.2.2.7 Système d'interactions embarqué

Un logiciel a été développé afin d'interagir à distance avec les programmes et données manipulés par le système d'acquisition au travers de plusieurs fonctionnalités. Avec cette application, il est possible de modifier le programme embarqué, de consulter et de modifier les paramètres expérimentaux, ainsi que de récupérer des données stockées. Les protocoles de communication de données aux standards Web (http, ftp) sont utilisés pour remplir ces fonctionnalités. Des webservices (http) permettent de contrôler l'expérimentation (acquisition en cours, temps restant, démarrage / arrêt d'une acquisition, etc.) ou d'obtenir des informations sur le stockage. Les transferts ftp permettent de télécharger les enregistrements disponibles. Pour faciliter le transfert de données, il a été nécessaire de transformer les données brutes (en ampères), qui nécessitaient des octets de données à double précision (64 bits) en valeurs entières plus légères (16 bits) en les convertissant en micro-ampères. Par conséquent, les données stockées par le système d'acquisition de données varient en réalité de 4000 μA à 20000 μA . Afin d'alléger encore plus les données pour optimiser les transferts, les données sont encodées en hexadécimal. La figure 3.13 schématise ce système.

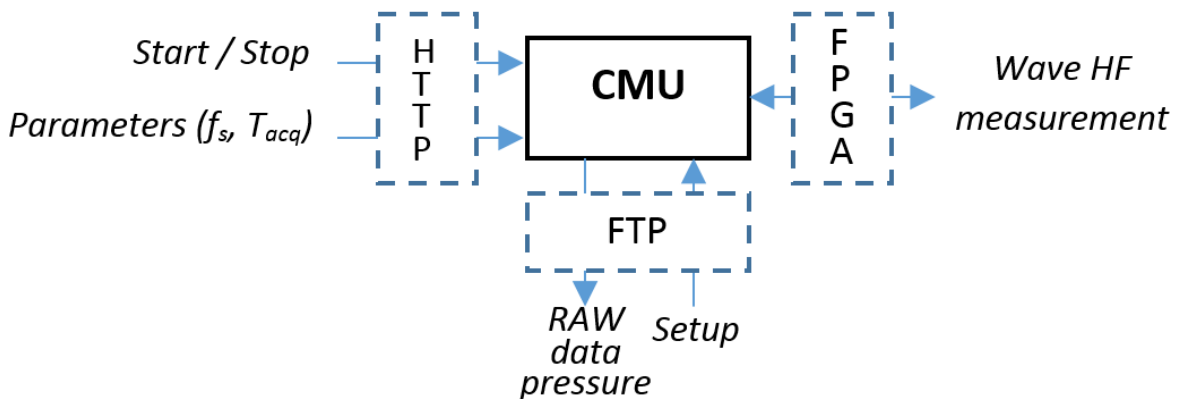


FIGURE 3.13 Schéma synthétisant le logiciel embarqué permettant de communiquer à distance avec la station. CMU : *Central Measurement Unit* ; F_s : fréquence d'échantillonnage ; T_{acq} : durée d'une acquisition.

3.2.2.8 Station dans son ensemble

Les conditions météorologiques étant extrêmes en période de tempête, l'équipement d'instrumentation a été choisi en conséquence pour sa robustesse, sa flexibilité et sa faible consommation énergétique. La station est contrôlable à distance permettant d'y avoir accès sans se rendre sur site. Par défaut, la station fonctionne en mode *burst*, c'est-à-dire qu'elle enregistre les données de pressions à intervalle régulier pendant une durée T_{acq} paramétrée. En effet, la station n'enregistre pas en permanence, mais plutôt 10 minutes de données à

chaque début d'heure, 24h/24, 7j/7. Cela permet de limiter le nombre de données tout en permettant d'avoir des données de manière régulière et constante. Grâce à cela nous pouvons constituer un échantillon de données suffisamment large pour avoir des pressions d'impact dans toutes les configurations météorologiques possibles. Bien que la fréquence normale des capteurs que nous utilisons soit de 1kHz, la station enregistre avec une fréquence d'échantillonnage $F_e = 10kHz$. En effet, le théorème de Shannon [81] affirme que la représentation discrète d'un signal exige des échantillons régulièrement espacés à une fréquence d'échantillonnage supérieure au double de la fréquence maximale présente dans ce signal. Il est donc nécessaire de suréchantillonner, et nous avons choisi une fréquence d'échantillonnage 10 fois supérieure.

Malgré l'environnement hostile, la station fonctionne en permanence et est en général mise à l'arrêt seulement en été afin de faire de la maintenance. En effet, en été il y a moins de tempêtes ce qui rend cette période idéale pour arrêter la station afin de vérifier que tout fonctionne correctement. Nous avons par exemple eu à changer de batterie car la première mise en place a commencé à perdre de la charge. Nous avons également changé le router car le précédent chauffait trop.

La figure 3.14 schématise la station dans son ensemble. Sur cette figure nous pouvons voir les trois parties principales, qui sont l'installation électrique, l'acquisition des données des capteurs de pression et le système de transmission. La photo 3.15 montre l'installation électrique.

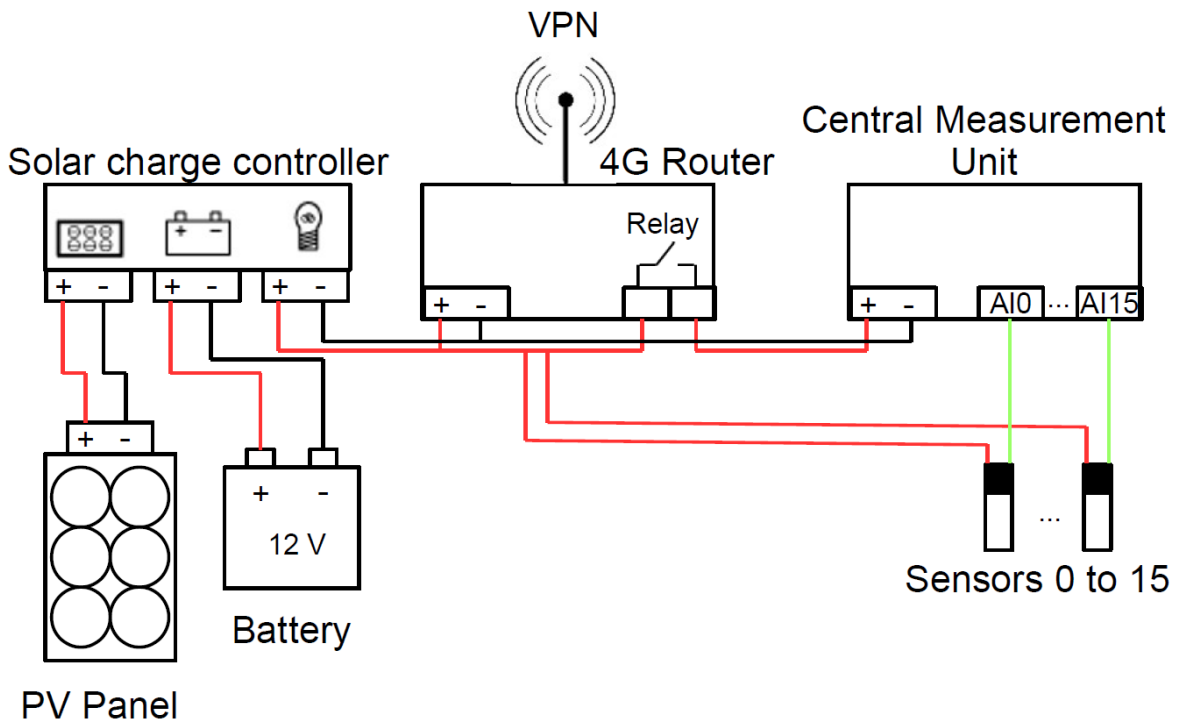


FIGURE 3.14 Schéma de la configuration de l'instrumentation pour l'acquisition et la transmission de données autonomes (CMU : Central Measurement Unit ; VPN : Virtual Private Network ; PV : panneau photovoltaïque).



FIGURE 3.15 Photo du panneau solaire et de l'installation électrique intégrée dans la digue.

3.2.2.9 Mise en place de la station sur la digue de l'Artha

Cette section a pour simple objectif de résumer comment nous avons installé la station de la digue de l'Artha. Des photos illustrant ceci sont disponibles en annexe C. L'installation a nécessité de gros travaux sur la digue. Ces travaux ont été réalisés par le Conseil Départemental 64 (CD64) [82].

Il a fallu faire une saignée d'environ 50m sur la digue pour installer une gaine entre le musoir ouest où sont positionnés les capteurs et l'arrière de la digue où se trouve la station. La photo C.1 montre un agent du CD64 en train de réaliser la saignée.

En parallèle, d'autres agents remettaient en état l'endroit abrité à l'arrière de la digue, là où la station sera placée. La figure C.2 montre cet emplacement. Cet emplacement a été abimé à cause des conditions extrêmes. En effet, les parois à l'intérieur commençaient à s'effriter, quelques fissures étaient présentes. Il a donc fallu rattraper ces dégâts afin d'éviter que l'état n'empire dans le futur et que la station puisse être à l'abri.

Dans le même temps, sur le musoir ouest des agents ont finalisés la mise en place des caissons qui vont recevoir les plaques sur lesquelles les capteurs de pression sont fixés. La photo C.3 montre cela.

Il a fallu installer tous les capteurs sur les deux plaques qui constituent la croix. La photo C.4 montre les capteurs placés sur les deux plaques ainsi que tous les câbles enroulés. Ces câbles seront tirés dans une gaine qui reliera l'emplacement des capteurs à la station.

Ensuite, la gaine a pu être mise en place, et les câbles des capteurs ont pu être tirés dans cette gaine. La photo C.5 est une photo de la mise en place de cette gaine. Nous pouvons également voir des agents en train de tirer les câbles des capteurs dans la gaine.

Une fois les câbles tirés et la gaine bien placée dans la saignée, il a fallu bétonner afin de consolider l'installation. La photo C.6 montre cette étape.

Enfin, la dernière étape a été de fixer les plaques sur les caissons du musoir ouest. Les photos C.7, C.8 et C.9 montrent la mise en place des deux plaques sur lesquelles les capteurs sont positionnés.

3.2.2.10 Données de pression de la station de l'Artha

La station enregistre donc les données provenant de ces 16 capteurs ($C = 16$), à une fréquence d'échantillonnage $Fe = 10kHz$ (ou $10.24kHz$ en fonction du paramétrage). Une acquisition de données de durée $D = 10$ minutes est réalisée à chaque début d'heure. Un fichier d'acquisition de données représente donc 96 000 000 valeurs de pression ($C \times Fe \times D \times 60$). Afin d'optimiser l'espace et les performances, les données de la station sont encodées sur 16 bits. La station fonctionnant en permanence, la station fait donc 24 acquisitions de 10 minutes par jour, ce qui équivaut à 168 acquisitions par semaine. Chaque fichier d'acquisition, compressé in-situ au format ZIP, a une taille de 121 Mo.

Cela représente un nombre important d'acquisitions, dont certaines ne sont pas intéressantes car elles ne contiennent pas d'impacts. En effet, lors des marées basses par exemple, ce qui arrive 2 fois par jour, le niveau d'eau est souvent trop bas et les vagues sont arrêtées par la barrière de blocs en béton qui protège la digue. Il n'y a donc pas d'intérêt pour nous à récupérer ces fichiers vides d'impacts, d'autant plus que cela consomme de l'énergie, du stockage et du temps de transfert. De ce fait, nous avons mis en place sur la digue un seuil de détection embarqué. Ce seuil, fixé à 0.0075A, est juste supérieur à la pression atmosphérique. Grâce à la formule présentée dans la section 3.2.2.1, nous pouvons calculer l'équivalent en bars :

$$5(7500 - 4000)/(20000 - 4000) = 1.09375 \text{ bar}$$

Il permet de vérifier s'il y a potentiellement eu des impacts sur le capteur 0 (capteur du bas), qui est celui qui est touché par toutes les vagues qui percutent la digue. Si dans les données enregistrées sur le capteur 0 il y a des valeurs supérieures à ce seuil, alors l'enregistrement est sauvegardé afin d'être rapatrié, sinon il est simplement supprimé.

3.2.3 Données provenant d'autres stations de mesure

Pour nos travaux, nous avons utilisé des données qui proviennent d'autres stations. En effet, dans la partie 4.7, nous avons utilisé les réseaux de neurones afin d'essayer de prédire la pression d'impact des vagues sur la digue de l'Artha en utilisant des données d'autres stations. Ces données sont des données environnementales qui proviennent des stations suivantes :

- La station météo de Socoa,
- Le marégraphe de Socoa,
- Le houlographe 06403 - Saint-Jean-de-Luz (situé à 1km au large de la digue de l'Artha).

La figure 3.16 montre la digue de l'Artha ainsi que les stations utilisées. Les sous-sections suivantes présentent les données pour chacune de ces stations.

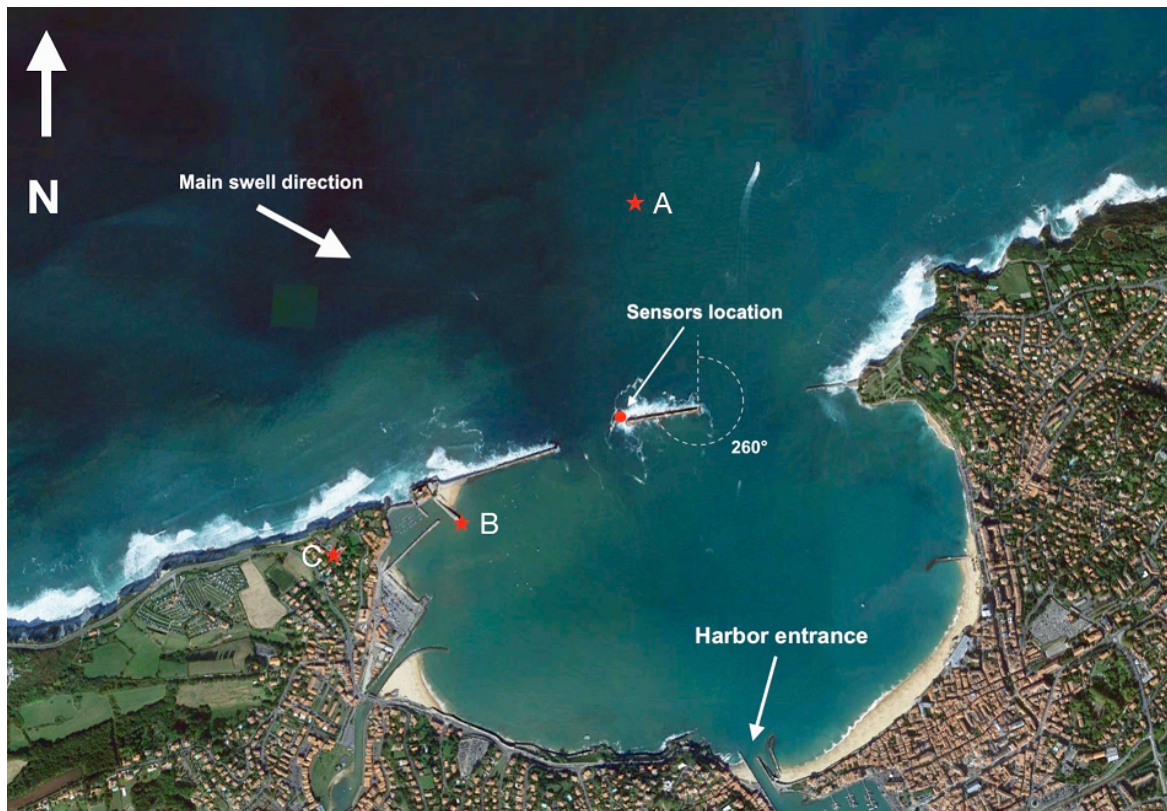


FIGURE 3.16 Digue de l'Artha avec l'emplacement des capteurs indiqué par un point rouge. Les étoiles rouges correspondent à (A) Houlographe 06403 ; (B) Marégraphe de Socoa ; (C) Station météo de Socoa [83].

3.2.3.1 Données de la station météo de Socoa

Ces données proviennent du site :

http://rp5.am/Archives_m%C3%A9t%C3%A9o_%C3%A0_Socoa [84].

Les figures A.1 et A.2 en annexe montrent un extrait de ces données.

Détails de chaque colonne :

- Heure locale à Socoa : Heure de la localité,
- T : Température de l'air ($^{\circ}\text{C}$) à 2 m au-dessus de la surface du sol,
- P_0 : Pression atmosphérique au niveau de la station (millimètres de mercure),
- P : Pression atmosphérique rapportée au niveau moyen de la mer (millimètres de mercure),
- P_a : Tendence barométrique : variation de la pression atmosphérique au cours des trois dernières heures (millimètres de mercure),
- U : Humidité relative de l'air (%) à 2 m au-dessus de la surface du sol,

- *DD* : Direction moyenne du vent (rose des vents) à 10-12 mètres au-dessus de la surface du sol sur la période de 10 minutes précédant immédiatement le moment de l'observation,
- *Ff* : Vitesse moyenne du vent à 10-12 mètres au-dessus de la surface du sol sur la période de 10 minutes précédant immédiatement le moment de l'observation (mètres par seconde),
- *ff10* : Valeur de la rafale maximale à 10-12 mètres au-dessus de la surface du sol sur la période de 10 minutes précédant immédiatement le moment de l'observation (mètres par seconde),
- *ff3* : Valeur de la rafale maximale à 10-12 mètres au-dessus de la surface du sol entre les périodes d'observation (mètres par seconde),
- *N* : Nébulosité totale,
- *WW* : Temps présent transmis par une station météorologique,
- *W1* : Temps antérieur entre les périodes d'observation 1,
- *W2* : Temps antérieur entre les périodes d'observation 2,
- *Tn* : Température minimale de l'air (degrés Celsius) pour la période précédente (pas plus de 12 heures),
- *Tx* : Température maximale de l'air (degrés Celsius) pour la période précédente (pas plus de 12 heures),
- *Cl* : Nuages des genres Stratocumulus, Stratus, Cumulus et Cumulonimbus,
- *Nh* : Étendue de tous les nuages *Cl* présents ou, en l'absence de nuages *Cl*, étendue de tous les nuages *Cm* présents,
- *H* : Base du nuage le plus bas (m),
- *Cm* : Nuages des genres Altocumulus, Altostratus et Nimbostratus,
- *Ch* : Nuages des genres Cirrus, Cirrocumulus et Cirrostratus,
- *VV* : Distance horizontale de visibilité (km),
- *Td* : Température du point de rosée à 2 m au-dessus de la surface du sol (degrés Celsius),
- *RRR* : Total des précipitations (mm),
- *tR* : Période d'accumulation des précipitations,
- *E* : État du sol sans couche de neige ou de glace mesurable,
- *Tg* : Température minimale de la surface du sol pendant la nuit (degrés Celsius),
- *E'* : État du sol recouvert d'une couche de neige ou de glace mesurable,
- *sss* : Hauteur de la couche de neige, en centimètres.

3.2.3.2 Données du marégraphe de Socoa

Ces données proviennent du site :

<https://data.shom.fr/donnees/refmar/95/download#001=eyJjIjpbLTE4OTcxMS4yOTIyODk4ODYyLD>
[85].

Les figures A.3 en annexe montre un extrait de ces données.

Détails de chaque colonne :

- *Date* : Date et heure,
- *Valeur* : Niveau de l'eau en mètre,
- *Source* : Source de la donnée. Il existe 6 sources différentes :
 1. Données brutes temps réel,
 2. Données brutes temps différé,
 3. Données validées temps différé,
 4. Données horaires validés,
 5. Données horaires bruts,
 6. Pleines et basses mers.

3.2.3.3 Données du houlographe

Ces données proviennent du site :

<http://candhis.cetmef.developpement-durable.gouv.fr/campagne/?idcampagne=6c8349cc7260ae62e3b>
[86].

Les figures A.4 et A.5 en annexe montrent un extrait de ces données.

Détails de chaque colonne :

- *dateheure* : Horodate de l'enregistrement,
- *h13d* : Hauteur significative des vagues définies par passage au niveau moyen par valeurs décroissantes,
- *h110d* : Hauteur moyenne du dixième supérieur en hauteur des vagues définies par passage au niveau moyen par valeurs décroissantes,
- *hmaxd* : Hauteur maximale des vagues définies par passage au niveau moyen par valeurs décroissantes (front montant),
- *hsigma* : Hauteur significative des vagues estimée,

- *hrmsd* : Hauteur quadratique moyenne des vagues définies par passage au niveau moyen par valeurs décroissantes,
- *h2percentd* : Hauteur de vague dépassée par 2% de l'ensemble des vagues définies par passage au niveau moyen par valeurs décroissantes,
- *th13d* : Période significative des vagues définies par passage au niveau moyen par valeurs décroissantes,
- *th110d* : Moyenne des périodes du dixième supérieur en hauteur des vagues définies par passage au niveau moyen par valeurs décroissantes,
- *tavd* : Période moyenne des vagues définies par passage au niveau moyen par valeurs décroissantes,
- *thmaxd* : Période de la vague de hauteur maximale, les vagues étant définies par passage au niveau moyen par valeurs décroissantes,
- *tmaxd* : Période maximale des vagues définies par passage au niveau moyen par valeurs décroissantes,
- *etamax* : Élévation maximale de la surface libre par rapport au niveau moyen,
- *etamin* : Élévation minimale de la surface libre par rapport au niveau moyen,
- *sz13d* : Cambrure significative des vagues,
- *szmaxd* : Cambrure maximale des vagues,
- *tszmaxd* : Période de la vague de cambrure maximale,
- *nbrevag* : Nombre de vagues définies par passage au niveau moyen par valeurs décroissantes,
- *skew* : Asymétrie de l'élévation de la surface libre,
- *kurt* : Kurtosis de l'élévation de la surface libre,
- *rhh* : Coefficient de corrélation entre les hauteurs de vagues successives définies par passage au niveau moyen par valeurs décroissantes,
- *hm0* : Hauteur significative spectrale des vagues,
- *tp* : Période de pic barycentrique,
- *t02* : Période moyenne,
- *te* : Période moyenne énergétique,
- *eps2* : Indice d'étroitesse spectrale,
- *kapa* : Largeur spectrale,
- *thetap* : Direction de provenance au pic, mesurée par rapport au nord géographique, positivement dans le sens des aiguilles d'une montre,
- *thetam* : Direction de provenance moyennée sur le spectre, mesurée par rapport au nord géographique, positivement dans le sens des aiguilles d'une montre,
- *sigmap* : Largeur directionnelle au pic,

— *sigmam* : Largeur directionnelle moyenne.

3.3 Bilan

Dans ce chapitre nous avons détaillé la station de mesure que nous avons mise en place sur la digue de l'Artha, permettant d'enregistrer les pressions d'impact de vagues à haute fréquence sur cette structure côtière. 16 capteurs de pression permettent d'avoir une représentation 2D des pressions d'impact sur le mur de la digue, au niveau du musoir ouest.

Cette station a été installée afin de répondre aux besoins que nous avons exprimé afin de pouvoir étudier les pressions d'impact de vagues sur une structure en situation réelle.

Les données enregistrées par cette station vont être utilisées en combinaison avec des données provenant d'autres stations dans l'objectif de mieux comprendre le phénomène complexe que nous étudions.

Chapitre 4

Résultats

Dans les sections précédentes, nous avons vu que nous avons mis en place une station permettant de récolter un volume conséquent de données sur les pressions d'impact sur la digue de l'Artha. Nous avons également vu que des algorithmes de *data mining* et de *machine learning* peuvent aider à extraire de l'information de toutes ces données. Ce chapitre a donc pour objectif de présenter le cheminement de notre réflexion et de notre travail, ainsi que les résultats que nous avons obtenus.

Dans cette section nous allons commencer par présenter le prétraitement que nous avons appliqué aux enregistrements des pressions d'impact de la digue de l'Artha.

Ensuite nous allons expliquer l'algorithme de détection automatique d'impacts que nous avons mis en place afin d'extraire les impacts des enregistrements.

Après cela, nous présenterons les différents critères que nous calculons pour les impacts, ainsi que les bases de données qui en découlent.

Par la suite, nous présenterons le calcul de covariance permettant de connaître le délai d'impact entre deux capteurs.

Nous enchaînerons par la visualisation animée d'impact en vidéo de deux manières différentes : sur un axe vertical et en 2 dimensions.

Nous présenterons ensuite le travail que nous avons effectué afin d'essayer de faire de la prévision de puissance d'impact.

Nous finirons par une étude de clustering des impacts grâce à des algorithmes non-supervisés.

4.1 Prétraitement des signaux de la station de l'Artha

La première chose à faire avant d'utiliser un signal est de le prétraiter. Cela permet de supprimer le bruit qui peut s'introduire dans les signaux. Le bruit peut provenir soit de l'environnement propre, soit au cours de la chaîne de transmission et de traitement de l'information. L'objectif est de ne récupérer que le signal dont l'information est pertinente et utilisable.

Comme nous l'avons présenté dans l'article de revue [75], il y a trois étapes importantes pour prétraiter les données :

1. Conversion des données en bar,
2. Application d'un filtre non-linéaire,
3. Application d'un filtre linéaire.

4.1.1 Signaux bruts

Lors de l'acquisition, les signaux bruts sont enregistrés en μA . Un fichier de données, pesant environ 124 Mo, contient 10 minutes de données par capteur, avec un nombre de capteur égale à 16. Comme expliqué dans la section 3.2.2.10, cela représente 6 000 000 de valeurs de pression par capteur, soit 96 000 000 pour tous les capteurs. Pour la période de juin 2018 à juin 2019, la station a enregistré 3241 fichiers contenant des impacts. La figure 4.1 montre le signal brut en μA d'un capteur, sans aucun prétraitement.

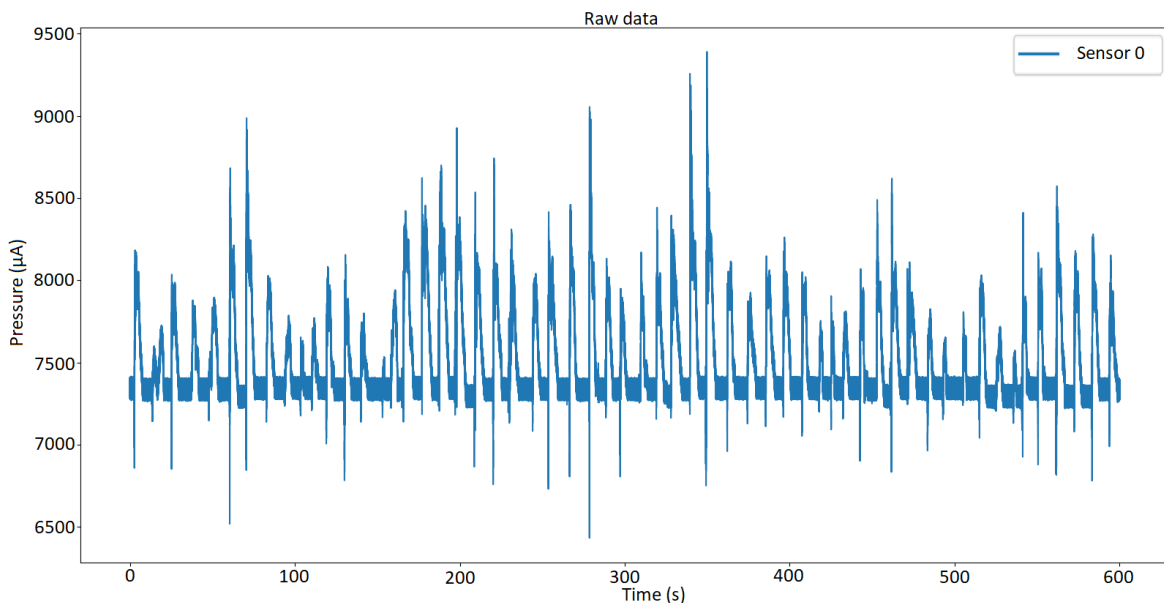


FIGURE 4.1 Signal brut en μA d'un capteur.

4.1.2 Conversion des données

La première étape est la conversion des données de μA en bars. En effet, la sortie électrique des capteurs varie entre 4 mA et 20 mA pour une pression qui varie de 0 à 5 bars, pour les capteurs Keller utilisés. La conversion est faite grâce à la formule présentée dans la section 3.2.2.1. La figure 4.2 montre le même signal que la figure précédente 4.1 mais où les données sont converties en bar.

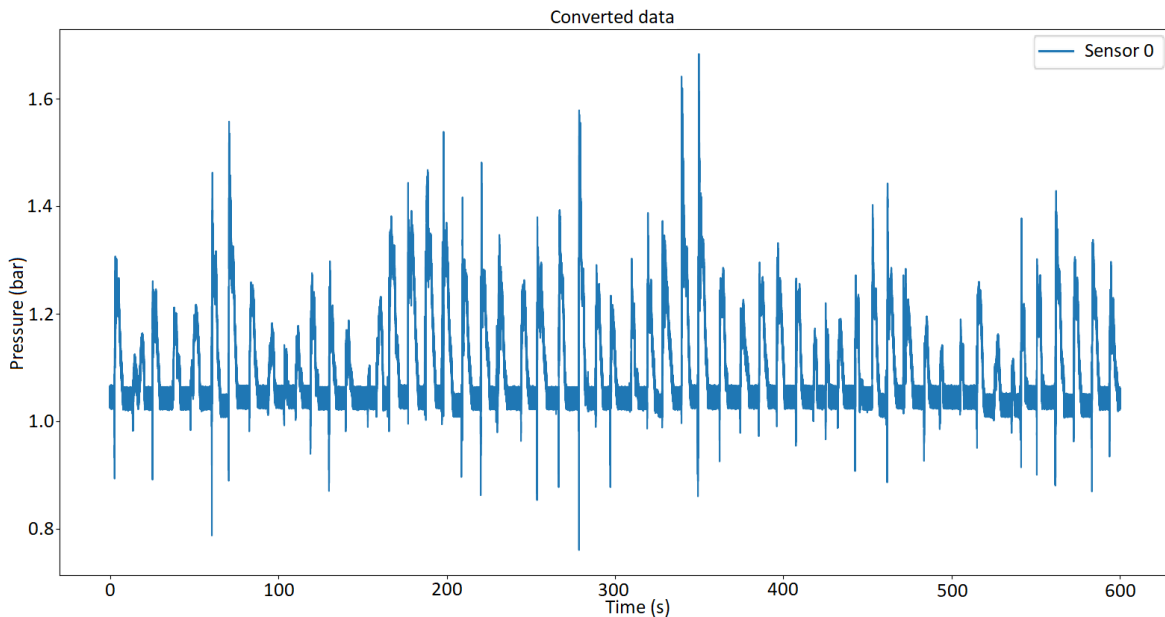


FIGURE 4.2 Signal converti en bar d'un capteur.

4.1.3 Filtrage non-linéaire

La deuxième étape est l'application d'un filtre non-linéaire. Cette catégorie de filtre est utilisée afin de réduire le bruit qui peut s'introduire dans un signal. Il permet notamment de corriger les singularités (valeurs aberrantes) qui peuvent s'introduire dans le signal.

Nous avons choisi un filtre médian. Pour une valeur V_n du signal, ce filtre prend les T valeurs de chaque côté de V_n , et les trie par ordre croissant. Ensuite, la valeur V_n est remplacée par la valeur médiane des valeurs triées. Le filtre passe ensuite à la valeur suivante, V_{n+1} et répète ces étapes, jusqu'à ce que toutes les valeurs du signal soient passées par le filtre médian. T est un paramètre du filtre médian, que l'on appelle la largeur de fenêtre. Il s'agit du nombre de valeurs que l'on veut prendre à droite et à gauche de la valeur V_n .

Dans ce travail, nous utilisons un filtre médian avec une largeur de fenêtre de 5. Ce paramètre a été déterminé car nous avons $Fe = 10kHz$ et nos capteurs ont une bande passante

de 1 kHz, soit 10 fois plus petite. Il faut donc prendre un nombre d'échantillons proche de 10 (10 étant impossible car le nombre d'échantillons utilisés par le filtre médian est impair), nous avons donc pris 11 : 5 de chaque côté, plus l'échantillon V_n .

La figure 4.3 montre un signal avant l'application du filtre médian. Nous pouvons observer qu'avant l'application du filtre, il y a un pic à environs 40 bars. Cet échantillon est forcément une aberration introduite par un bruit, car les capteurs que nous utilisons permettent d'enregistrer une pression maximum de 5 bars. L'application du filtre médian permet de corriger cette erreur et de rendre le signal utilisable pour notre étude, comme nous pouvons le voir sur la figure 4.4.

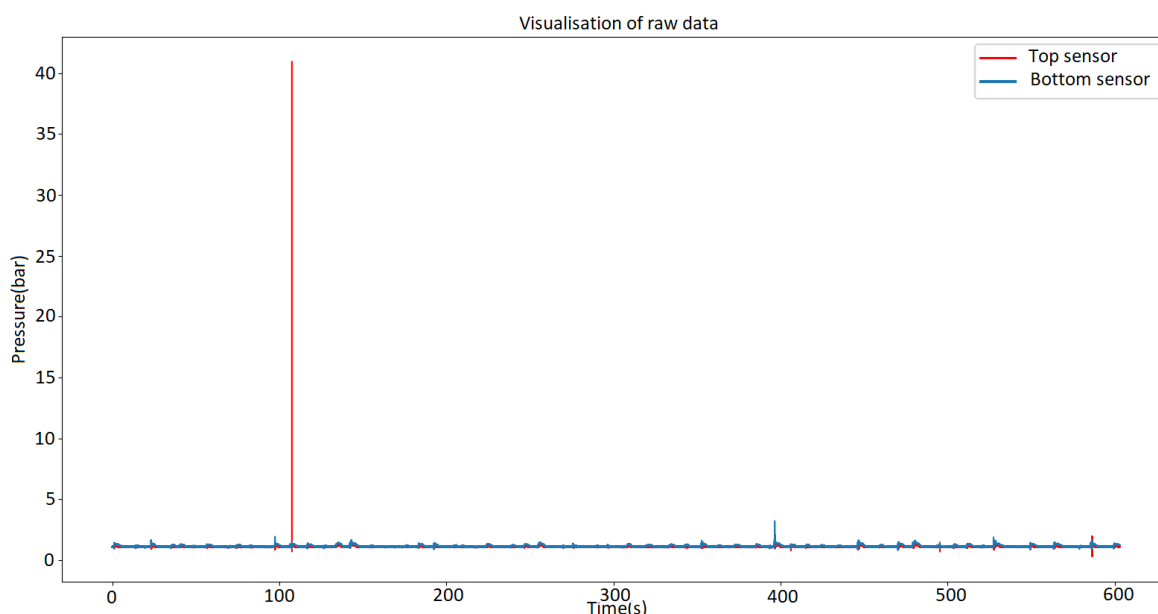


FIGURE 4.3 Signal avant application du filtre médian contenant une valeur aberrante.

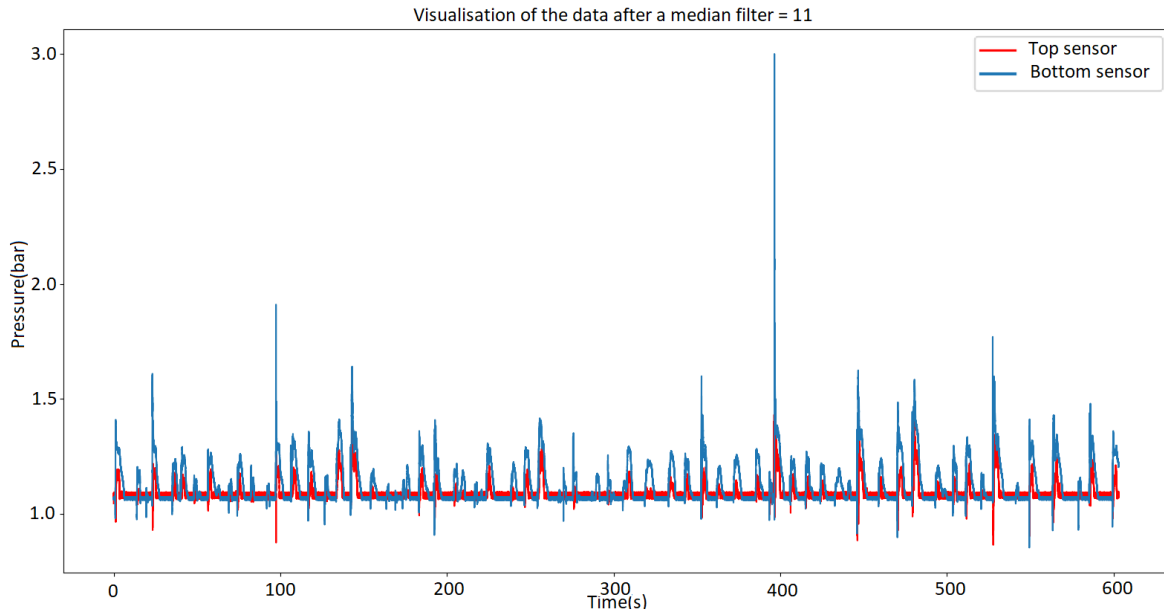


FIGURE 4.4 Signal après application du filtre médian.

4.1.4 Filtrage linéaire

La dernière étape de pré-traitement est l'application d'un filtre linéaire. Il existe 4 types de filtres linéaires :

- Passe-bas : Filtre toutes les fréquences supérieures à une fréquence maximale déterminée,
- Passe-haut : Filtre toutes les fréquences inférieures à une fréquence minimale donnée,
- Passe-bande : Filtre toutes les fréquences qui ne sont pas dans un intervalle déterminé par une fréquence minimale et une fréquence maximale,
- Rejecteur de bande : Filtre les fréquences d'une bande fréquentielle déterminée.

Dans notre cas, nous nous intéressons aux filtres passe-bas, car cela permet la suppression des fréquences supérieures à la bande passante des capteurs. En effet, la fréquence normale des capteurs que nous utilisons est de 1 kHz. Afin de suivre la loi de Shannon, nous avons suréchantillonné, en enregistrant les données à une fréquence d'échantillonnage de 10 kHz. De ce fait, il est nécessaire, avant utilisation des données, de supprimer les fréquences supérieures à 1 kHz.

4.1.4.1 Filtre de Butterworth

Le filtre de Butterworth est un filtre linéaire qui est fait pour avoir une réponse fréquentielle dans sa bande passante la plus plate possible. Cependant, la pente de coupure est relativement faible comparée à d'autres filtres. Il suffit d'augmenter l'ordre afin d'obtenir une pente de coupure plus raide. Ce filtre a une réponse très plate dans la bande passante.

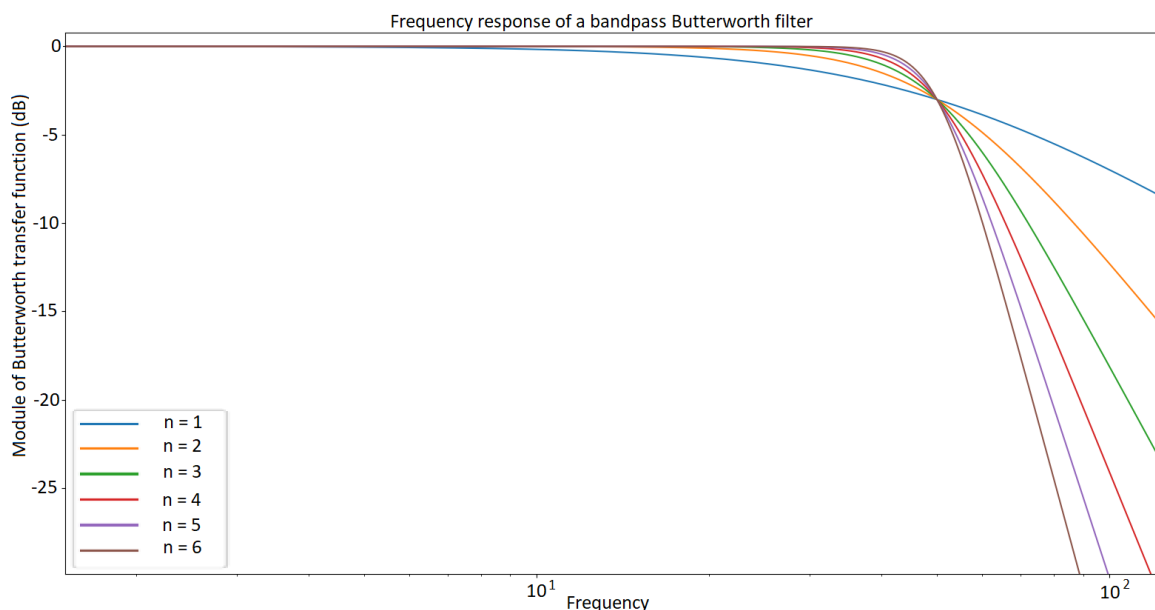


FIGURE 4.5 Réponse en fréquence du filtre de Butterworth.

La formule de ce filtre est la suivante :

$$|H_n(j\omega)| = \frac{1}{\sqrt{1 + (\omega/\omega_c)^{2n}}}$$

où :

- H_n est la fonction de transfert,
- ω est la fréquence angulaire (ou pulsation) du signal en radians par seconde,
- ω_c est la fréquence de coupure,
- n est l'ordre.

4.1.4.2 Filtre de Tchebychev

Le filtre de Tchebychev est un filtre linéaire qui est utilisé quand les spécifications du gabarit permettent une ondulation de la bande passante. La pente de coupure, à ordre identique, est plus raide que le filtre de Butterworth.

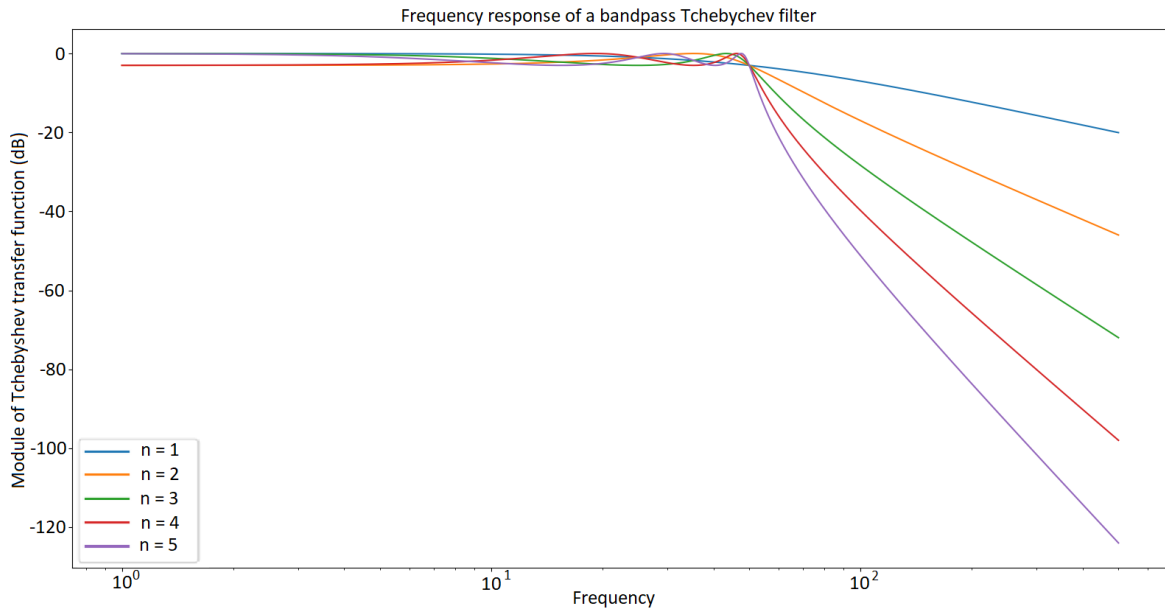


FIGURE 4.6 Réponse en fréquence du filtre de Tchebyshev.

La formule de ce filtre est la suivante :

$$|H_n(j\omega)| = \frac{1}{\sqrt{1 + \varepsilon^2 T_n^2\left(\frac{\omega}{\omega_0}\right)}}$$

où :

- H_n est la fonction de transfert,
- ε est le facteur d'ondulation,
- ω_c est la fréquence de coupure,
- T_n est le polynôme de Tchebyshev d'ordre n .

Sur les figures 4.5 et 4.6 nous voyons bien la différence entre les filtres de Butterworth et Tchebyshev. En effet, nous voyons que pour Butterworth, il n'y a pas de variations dans la bande passante, alors que pour Tchebyshev nous voyons bien les ondulations. En revanche, nous observons que pour le même ordre, la pente de coupure du filtre de Tchebyshev est plus raide.

Dans notre cas, nous avons besoin de ne laisser passer que les fréquences inférieures à 1 kHz. De ce fait, bien que pour un ordre donné le filtre de Tchebyshev coupe mieux, nous avons choisi d'utiliser le filtre de Butterworth car nous avons besoin d'une réponse fréquentielle la plus plate et précise possible dans la bande passante.

Nous avons donc utilisé un filtre Butterworth passe-bas, où nous avons fixé la fréquence de coupure à 1 kHz et l'ordre à 8. La figure 4.7 montre le résultat du prétraitement. Ce signal est le même que celui des figures 4.1 et 4.2.

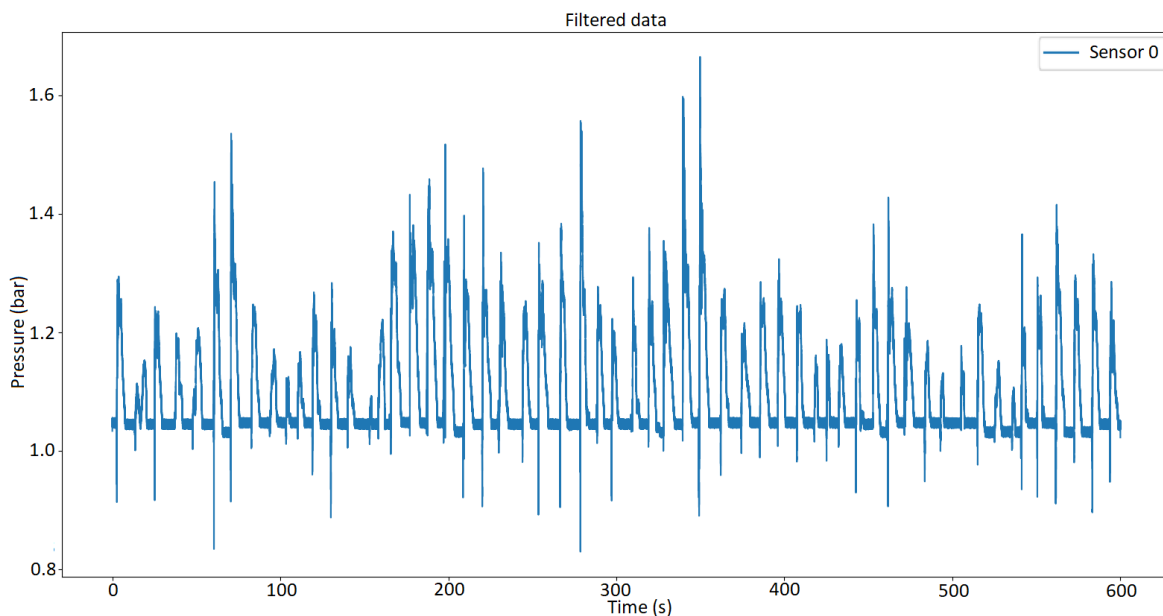


FIGURE 4.7 Signal filtré d'un capteur.

4.1.5 Temps de calcul

Sur un ordinateur possédant 32 Go de RAM et étant équipé d'un processeur Intel(R) Core(TM) i5-7500 CPU @ 3.40 GHz, l'ensemble de ce prétraitement demande environ 35 secondes de calcul pour un enregistrement de 10 minutes comprenant les signaux de 16 capteurs. Pour être plus précis, le temps de calcul est divisé comme suit par étape du prétraitement :

- Extraction de données et conversion de hexadécimal en decimal : 10.2 secondes,
- Conversion des données de μA en bar : 5.5 secondes,
- Application du filtre médian : 16.5 secondes,
- Application du filtre Butterworth 2.5 secondes.

Cet algorithme met donc au total environ 34.7 secondes pour extraire et prétraiter un fichier d'acquisition de 10 minutes contenant les signaux de 16 capteurs.

4.2 Détection automatique des impacts

4.2.1 Objectif

La station de la digue de l'Artha nous permet d'enregistrer des données pendant une période de 10 minutes à chaque début d'heure. Les capteurs se trouvant sur la paroi de la digue, à l'air libre, les signaux enregistrés correspondent à une alternance de pression atmosphérique et de pression d'impact. Cependant, ici seules les pressions d'impacts nous intéressent. En effet, pour l'objectif de notre travail, nous souhaitons obtenir de l'information sur chaque impact enregistré. Il faut donc être capable d'extraire les impacts qui ont eu lieu durant les 10 minutes d'enregistrement.

Jusqu'à présent, les données utilisées pour réaliser des études sur l'impact des vagues étaient analysées manuellement. Les fichiers de signaux devaient donc être ouverts un par un, et l'extraction d'une partie de signal pertinente était faite manuellement par un expert, comme cela avait été fait pour l'article [75]. L'avantage de cette approche est qu'elle permet de n'extraire que les informations qui sont pertinentes pour l'étude. Cependant il y a plusieurs inconvénients. Le premier est que cela est faisable lorsqu'il y a un volume restreint de données, mais lorsqu'il y en a une grosse masse, cela devient vite fastidieux. Le second inconvénient est que d'une étude à une autre, la pertinence des données peut évoluer. En effet, pour une étude, seules les données des impacts puissants vont être extraites. Mais, si une autre étude sur les mêmes données nécessite d'utiliser tous les impacts, il faudra alors retraiter tous les signaux pour extraire les informations nécessaires.

Il y a donc dans notre cas, un intérêt à traiter les données automatiquement afin de gagner du temps et d'extraire un maximum d'informations. Le but est de construire une base de données avec les informations extraites, ce qui permettra à un expert de sélectionner les données qui sont pertinentes pour son étude, sans avoir à ouvrir l'ensemble des enregistrements existants.

Cet algorithme est donc un point clé de notre contribution car c'est sur l'extraction des impacts dans les enregistrements que se base la suite du travail. Pour cette raison, nous avons beaucoup travaillé sur cet algorithme. En effet, il y a eu plusieurs versions dans lesquelles nous avons testé différentes approches. Il faut avant tout préciser que nous avons commencé en appliquant différents algorithmes de détection de pics. Cependant, ces algorithmes habituels de détection de pics, tels que les méthodes courantes utilisées dans MatLab Signal Processing Toolbox, ou tels que la transformée en ondelettes (Du, 2006 [87]), semblent échouer lorsqu'ils sont appliqués aux signaux d'impact des vagues enregistrés sur

la digue de l'Artha. En effet, ces signaux peuvent être non conventionnels, avec par exemple plusieurs pics dans un signal d'impact, comme le montre la figure 4.8.

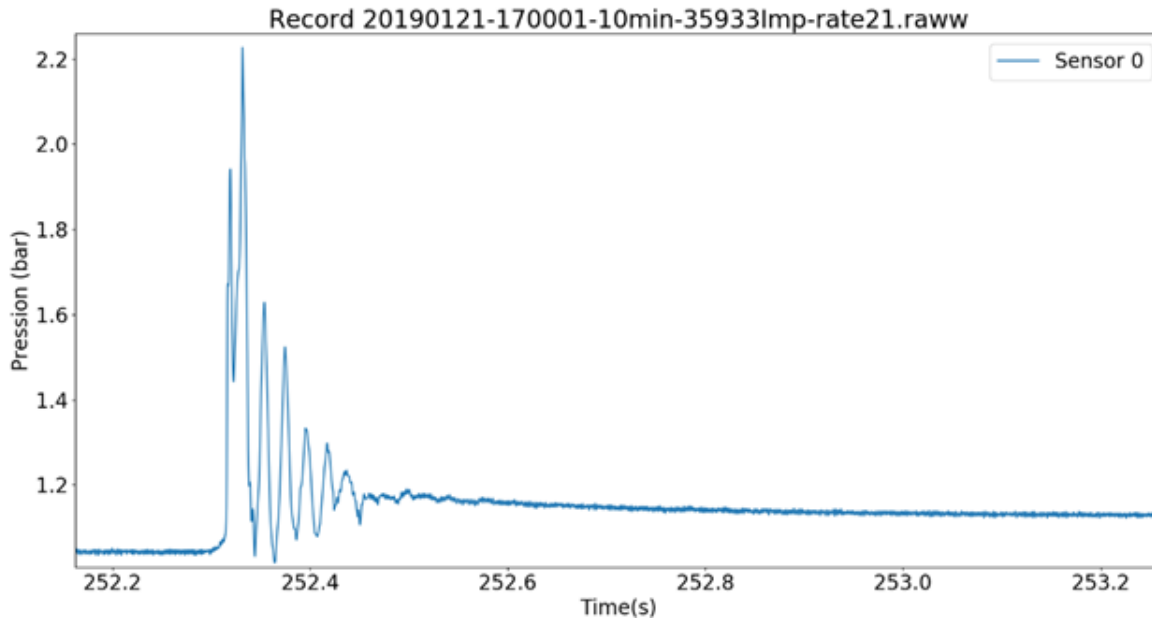


FIGURE 4.8 Signal d'impact contenant plusieurs pics.

De plus, ces signaux contiennent des variations de la composante continue, ce qui les rend encore plus difficiles à traiter. Pour pouvoir détecter efficacement les impacts dans les signaux, un algorithme spécifique a été implémenté. Les parties suivantes présentent donc les différentes approches que nous avons testées.

4.2.2 Algorithmes

Nous avons testé différentes approches pour le découpage des impacts. En effet, il n'y avait pas de solution triviale pour le découpage des impacts. Dans cette section, nous présentons les trois approches principales que nous avons testées.

4.2.2.1 Approche 1 : Algorithme sans paramètres empiriques

Cette première approche avait pour but ambitieux de détecter les impacts sans paramètres empiriques. L'objectif était d'avoir un algorithme utilisable en l'état, sans paramétrage, sur d'autres stations dans le cas où l'on souhaiterait réaliser de nouvelles expérimentations ailleurs.

Nous avons donc choisi de baser cet algorithme sur la moyenne du signal. En effet, la moyenne est utilisée comme seuil de détection des impacts. Cette solution fonctionne relativement bien, mais a quand même certains effets négatifs. Le principal est que lors de grosses tempêtes, la moyenne de l'enregistrement est élevée dû aux nombreux impacts de forte amplitude, et donc dans certains cas l'algorithme ne détecte pas les impacts plus petits dans ces signaux. Il y avait donc une perte d'information, que nous souhaitions éviter au maximum.

Lors des premiers résultats nous permettant de valider la qualité de l'algorithme, nous avons également constaté que dans certains cas, plutôt rares mais quand même existants, la composante continue évoluait pendant l'enregistrement. La figure 4.9 ci-dessous montre le changement de la composante continue que nous avons pu observer sur plusieurs enregistrements. La composante continue moyenne est représentée par le trait rouge, et nous pouvons voir dans l'encadré noir que la composante continue à cet instant précis est supérieure à la composante continue moyenne.

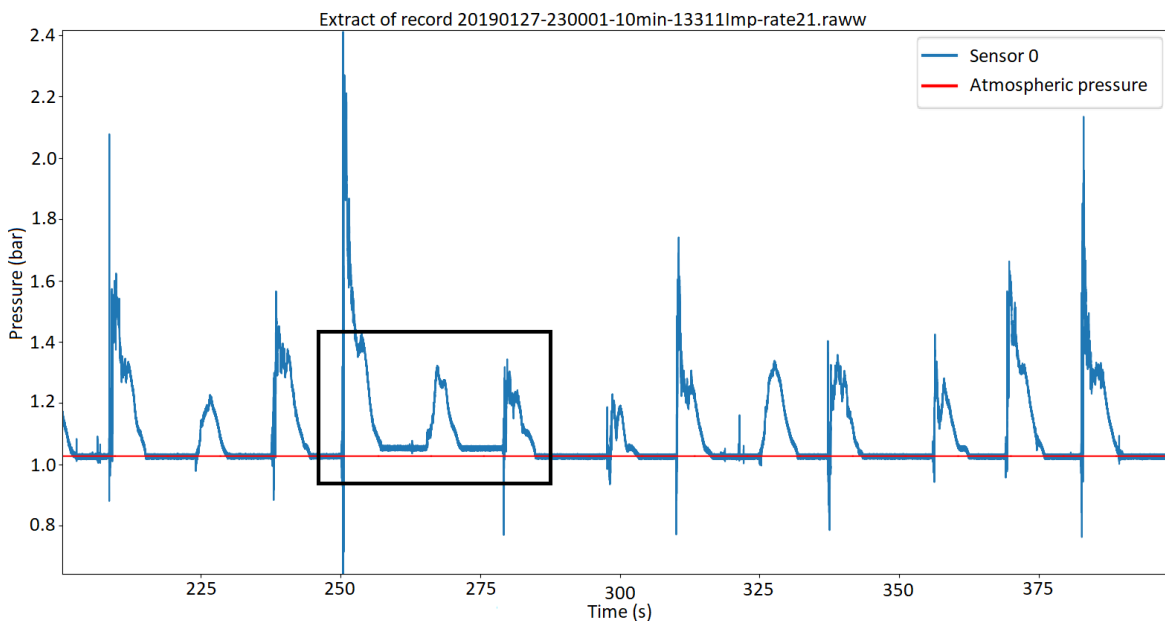


FIGURE 4.9 Signal contenant un décalage de la composante continue.

Malgré le fait qu'il soit mineur, ce changement inexplicable a posé problèmes, car cela impliquait que l'algorithme n'était pas capable de déterminer à quel moment l'impact commençait et finissait exactement. Pour ces raisons, nous avons donc conclu que cette approche n'était pas performante.

4.2.2.2 Approche 2 : Algorithme basé sur la dérivée

Cette seconde approche diffère beaucoup de la précédente. En effet, premièrement, cet algorithme utilise la dérivée du signal pour détecter les impacts. Deuxièmement, nous avons choisi d'utiliser des paramètres empiriques pour permettre de mieux détecter les impacts. Ces critères empiriques ont été déterminés après l'analyse manuelle d'une grosse masse de signaux de la digue de l'Artha. Le premier paramètre est un seuil qui permet de détecter le début d'un impact dans la dérivée. Ce seuil est fixé à 0.08 bar par seconde, car l'observation de la dérivée a permis de déterminer qu'un impact commençait lorsque la courbe franchissait ce seuil. De ce fait, si la dérivée du signal passe au-dessus de ce seuil, cela signifie que c'est le début d'un impact. Le deuxième paramètre permet de déterminer la fin de l'impact commencé. Nous avons déterminé qu'un impact prend fin quand la dérivée du signal croise 150 fois, positivement ou négativement, le zéro.

La figure 4.10 montre la dérivée d'une portion de signal. La dérivée du signal est en bleu. Le trait noir est le seuil positif de 0.08 bar par seconde. Le trait rouge est le seuil négatif de -0,08 bar par seconde. Enfin, le point rouge correspond au début de l'impact et le point noir à sa fin.

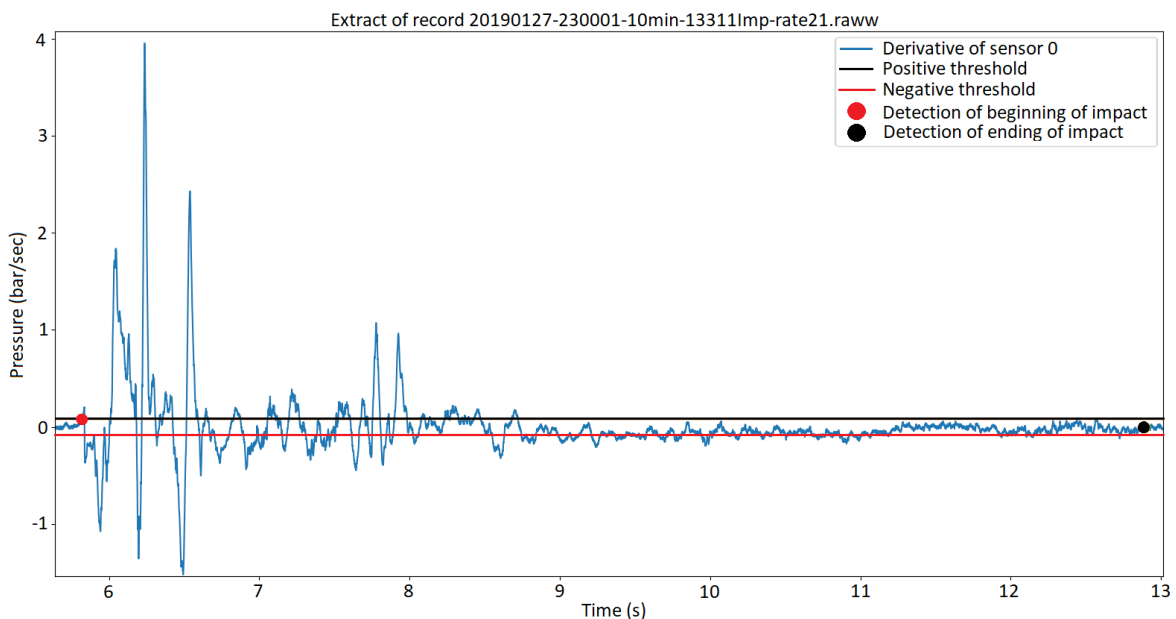


FIGURE 4.10 Dérivée d'une partie d'un signal avec seuils de détections.

Le début d'un impact est donc détecté quand la dérivée passe au-dessus du trait noir. Un impact commencé prend fin lorsque la dérivée croise 150 fois le zéro, sans repasser au-dessus du trait noir ou en dessous du trait rouge.

Cet algorithme fonctionne bien. Sur une centaine d'enregistrements testés, qui comprennent 16 signaux (capteurs) chacun, nous avons déterminé un taux d'erreur d'environ 3%, où la majorité des erreurs sont de faux positifs.

Cependant, en explorant les données et les résultats sur l'ensemble du jeu de données, nous nous sommes rendu compte que dans certaines conditions de tempête, l'algorithme était en erreur. C'est notamment le cas quand le vent souffle fort en direction de la digue. Les variations introduites dans le signal par le vent vont perturber l'algorithme et celui-ci va mal détecter les impacts.

De plus, le calcul de la dérivée est assez lourd au niveau calculatoire, ce que provoque un temps de calcul de l'algorithme trop long par rapport à ce que nous souhaitons avoir. Pour finir, un problème de cet algorithme était le second paramètre empirique, qui aurait dû être adapté dans le cas où l'on changerait la fréquence d'échantillonnage. En effet, le passage par zéro dans la dérivée fixé à 150 occurrences pour déterminer la fin d'un impact est directement lié à la fréquence d'échantillonnage du signal.

4.2.2.3 Approche 3 : Algorithme basé sur la suppression de la composante continue

Cette troisième et dernière approche est, finalement, la plus intuitive et la plus performante que nous ayons testée.

En effet, dans les deux approches précédentes, nous travaillions sur le signal absolu, donc avec une composante continue équivalente à la pression atmosphérique. Le problème est que cette composante continue varie d'un enregistrement à l'autre, d'un capteur à l'autre et même parfois, dans le temps pour un enregistrement donné sur un même capteur, comme le montre la figure 4.9. Il était donc impossible de fixer une valeur spécifique pour la composante continue. C'est pour cette raison que dans les deux algorithmes précédents nous avons travaillé avec le signal absolu. Une fois les impacts détectés, nous étions capables de déterminer la valeur de la pression atmosphérique, qui correspond à la moyenne du signal entre chaque impact. La figure 4.11 montre la pression atmosphérique dans les encadrés rouges.

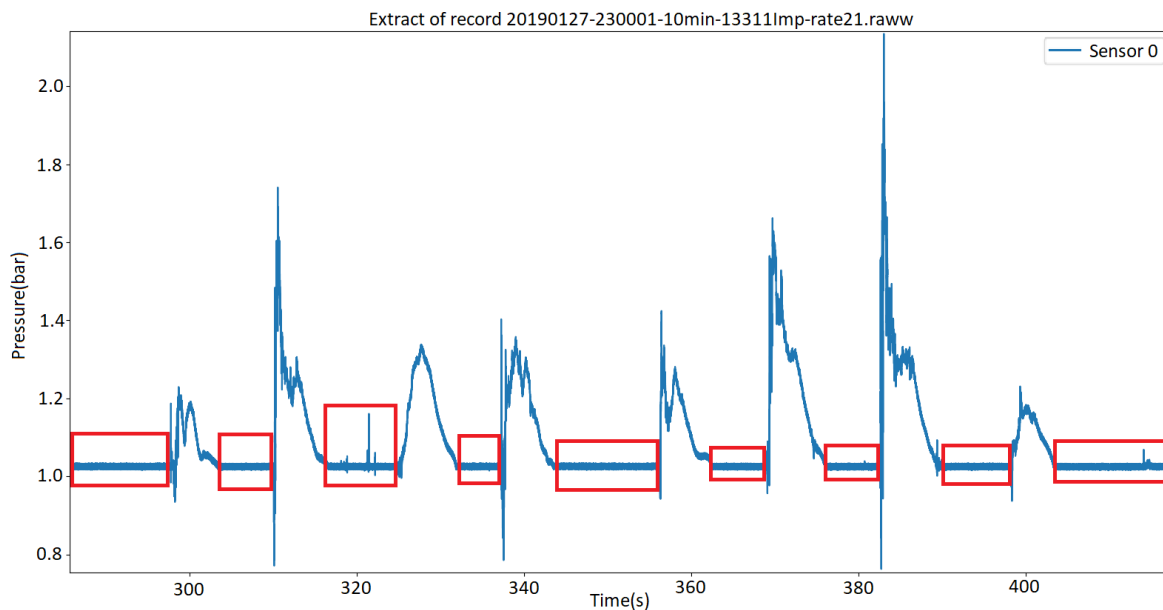


FIGURE 4.11 Pression atmosphérique entre chaque impact.

Evidemment, au tout début de notre travail, nous avons essayé de supprimer la composante continue. Une façon commune de faire cela est d'utiliser un filtre passe-haut. Mais comme le montre la figure 4.12, cela modifie le signal. Le signal d'impact avant le filtre passe-haut est en bleu et le signal après filtrage est en orange.

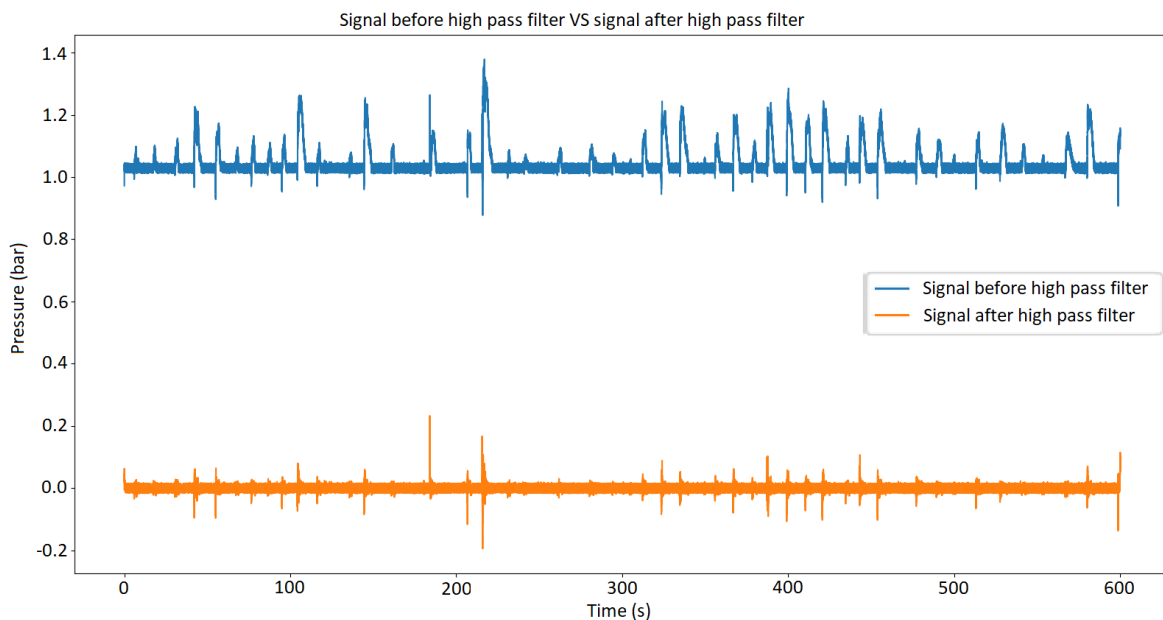


FIGURE 4.12 Signal avant et après application d'un filtre passe-haut.

Dans notre cas, nous ne souhaitons pas altérer le signal. Il a donc fallu trouver une autre manière permettant de supprimer la composante continue. Cette façon de faire est relativement simple. Il s'agit simplement de compter dans le signal entier, quelle est la valeur la plus présente. Cette valeur est forcément la pression atmosphérique. A partir de là, il est facile de retirer la pression statique du signal, et de travailler sur un signal relatif aligné à zéro.

Une fois ce traitement effectué, nous avons déterminé un seuil de détection empirique : ici 0.05 bar. La détection d'un impact commence donc quand le signal passe au-dessus de 0.05 bar et finit quand le signal passe en dessous des 0.05 bar. Une fois l'impact détecté, il suffit de redescendre de 0.05 à 0 avant et après l'impact afin d'avoir l'impact dans sa globalité, partant de 0 bar et finissant à 0 bar (relatif).

Cet algorithme a été testé sur 50 enregistrements, comprenant chacun les signaux de 16 capteurs. Nous avons choisi une dizaine d'enregistrements en conditions de tempête, et 40 pris au hasard parmi les 3241 enregistrements obtenus durant la période juin 2018 à juin 2019. L'objectif était de tester l'algorithme sur plusieurs combinaisons de conditions météorologiques différentes, notamment sur les périodes de tempêtes où l'approche précédente échouait. Le taux de réussite de cet algorithme est de 100%, vu que tous les impacts sont correctement détectés.

Seuls deux problèmes existent avec cette approche. Le premier se produit en période de tempêtes : entre deux impacts il arrive qu'il y ait des jets rapides sur les capteurs. Ces jets sont détectés comme étant des impacts lorsque leur pression dépasse 0.05 bar. Il s'agit donc de faux positifs, mais ce ne sont pas des erreurs bloquantes. Le deuxième problème reste la pression statique qui varie dans un signal, comme montré précédemment sur la figure 4.9. Cependant, avec cet algorithme, vu que nous savons que le signal doit être aligné sur 0, si nous n'arrivons pas à redescendre à 0, alors cela permet de détecter une variation de pression dans le signal et donc de relever l'anomalie. La problématique est donc toujours existante, mais détectée.

Cet algorithme met, sur un ordinateur possédant 32 Go de RAM et étant équipé d'un processeur Intel(R) Core(TM) i5-7500 CPU @ 3.40 GHz, environ 40 secondes pour détecter les impacts d'un enregistrement de 10 minutes comprenant les signaux de 16 capteurs.

4.2.2.4 Conclusion

L'algorithme de détection d'impact est un point clé de notre travail. En effet, toute la suite de nos travaux nécessite et découle des résultats de cet algorithme. C'est pour cette

raison que nous avons testé plusieurs approches, afin d'optimiser au mieux les résultats obtenus, que ce soit en termes de performance ou en termes de détection d'impacts.

Pour l'élaboration de cet algorithme, nous avons testé divers outils déjà existants qui n'ont pas donné de résultats satisfaisants. Nous avons donc réfléchi, implémenté et testé différentes approches, dont les trois principales ont été décrites dans cette section.

Les résultats obtenus nous ont fait choisir la troisième approche présentée, celle basée sur la suppression de la composante continue. En effet, cet algorithme est plus performant en termes de temps de calcul, mais aussi en termes de succès de détection d'impacts. C'est pour ces raisons que nous avons choisi de travailler avec cette approche.

4.3 Calcul de critères d'impact

Une fois que ces impacts sont détectés, il est important de pouvoir les caractériser. Pour cela, il est nécessaire de calculer différents critères pour chaque impact. Nous avons calculé de nombreux critères afin de caractériser un impact, mais nous n'en avons finalement retenus que 9 qui sont les plus importants et pertinents. Dans cette section nous allons présenter ces critères sélectionnés, qui peuvent être séparés en deux catégories : critères temporels et critères de pression.

4.3.1 Critères temporels

4.3.1.1 Temps de montée

Le temps de montée correspond à la durée que met le signal pour arriver à son maximum. Par convention, dans la discipline du traitement du signal, le temps de montée est calculé entre 10% du maximum et 90% du maximum. La figure 4.13 illustre ce critère.

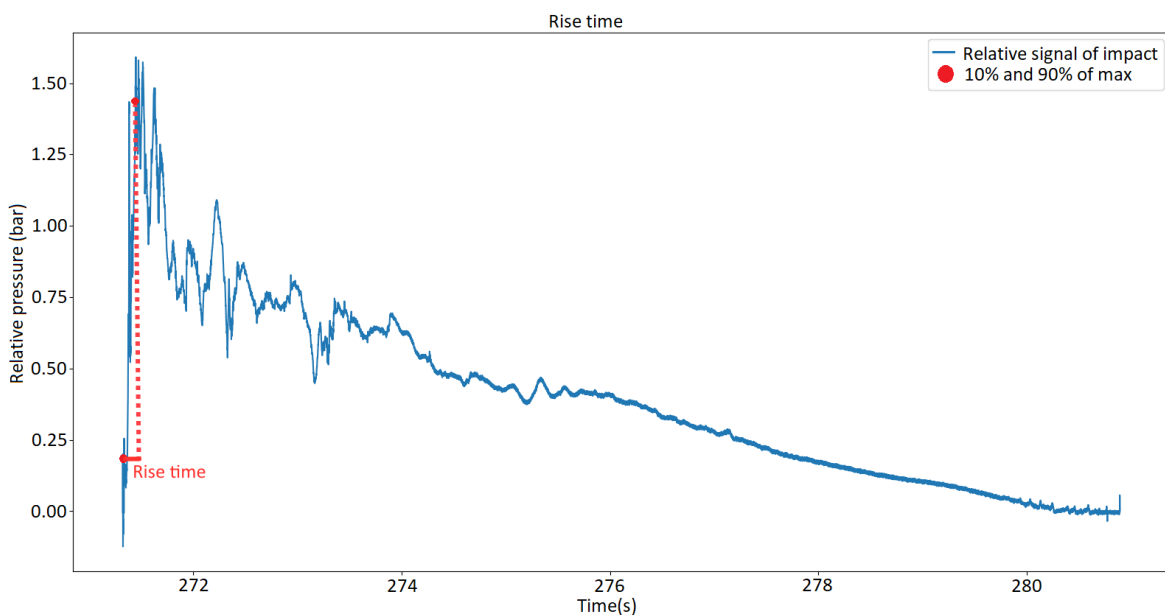


FIGURE 4.13 Temps de montée de 10% à 90% du maximum.

4.3.1.2 Temps de descente

Le temps de descente est la durée que met le signal pour redescendre à son minimum. Comme pour le temps de montée, il convient de le calculer entre 90% du maximum et 10% du maximum. La figure 4.14 illustre ce critère.

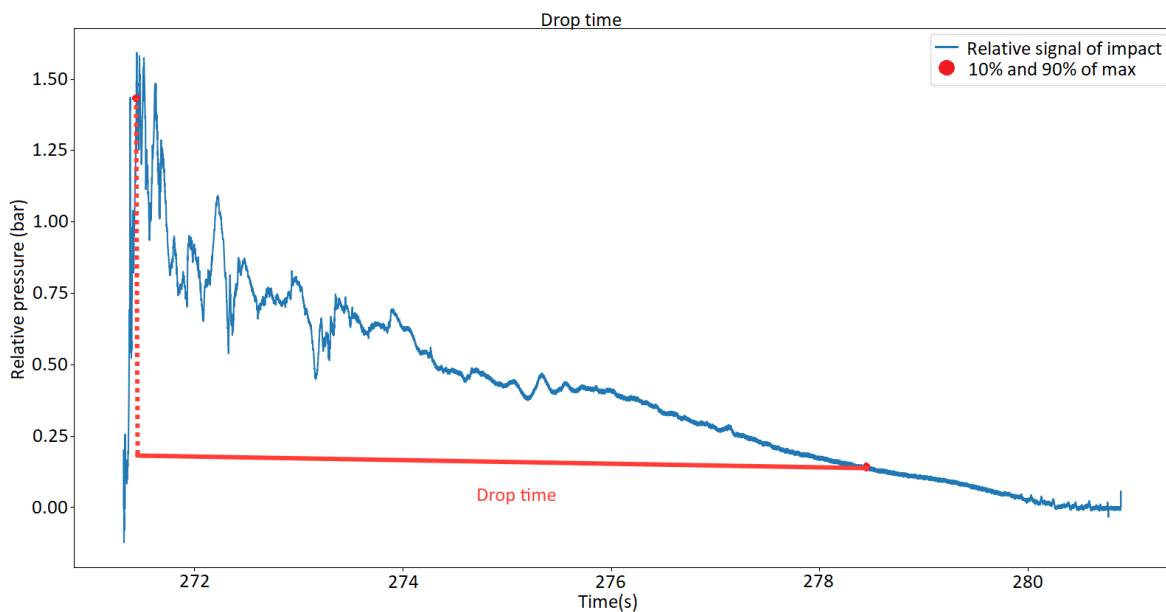


FIGURE 4.14 Temps de descente de 90% à 10% du maximum.

4.3.1.3 Durée

Il s'agit de la durée de l'impact. A l'instar du temps de montée et du temps de descente, la durée est calculée comme l'intégrale, de 10% du maximum avant le maximum à 10% du maximum après le maximum. La figure 4.15 illustre ce critère.

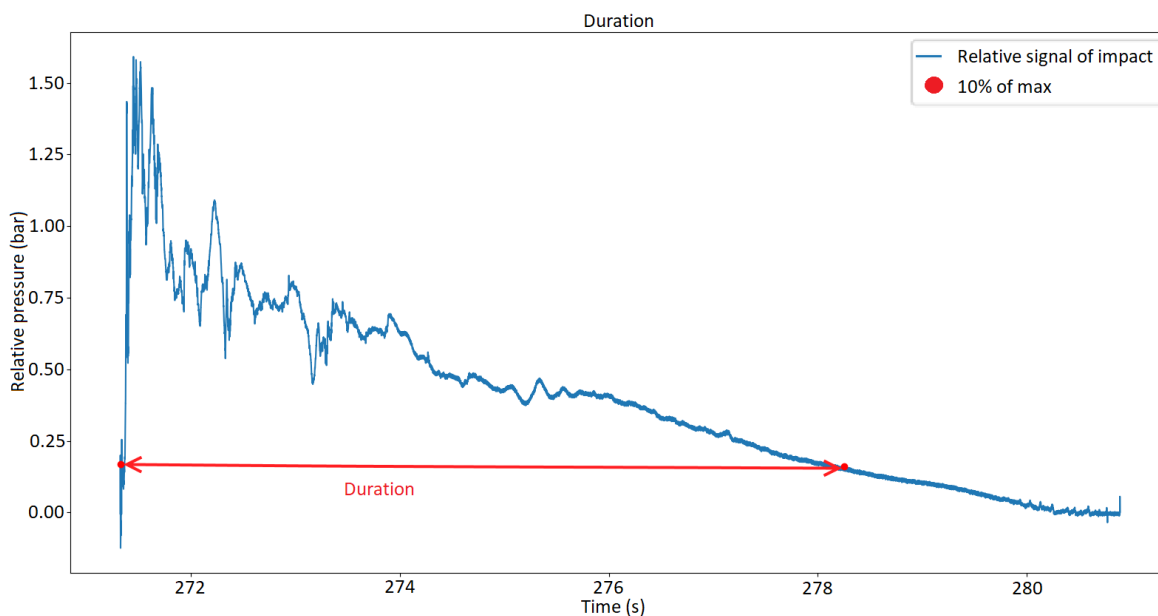


FIGURE 4.15 Durée de 10% du maximum avant le pic à 10% du maximum après le pic.

4.3.2 Critères de pression

4.3.2.1 Maximum

Le maximum correspond simplement à la valeur maximale enregistrée dans le signal d'impact. Pour ce critère, il existe deux valeurs différentes : la valeur absolue du maximum et la valeur relative du maximum. La valeur absolue correspond à la valeur telle qu'elle est enregistrée par la station, c'est-à-dire en incluant la pression atmosphérique. La valeur relative quant à elle correspond à la valeur absolue moins la pression atmosphérique. Elle permet donc de connaître la pression réelle engendrée par l'impact de la vague. La figure 4.16 montre ces deux valeurs pour le même impact.

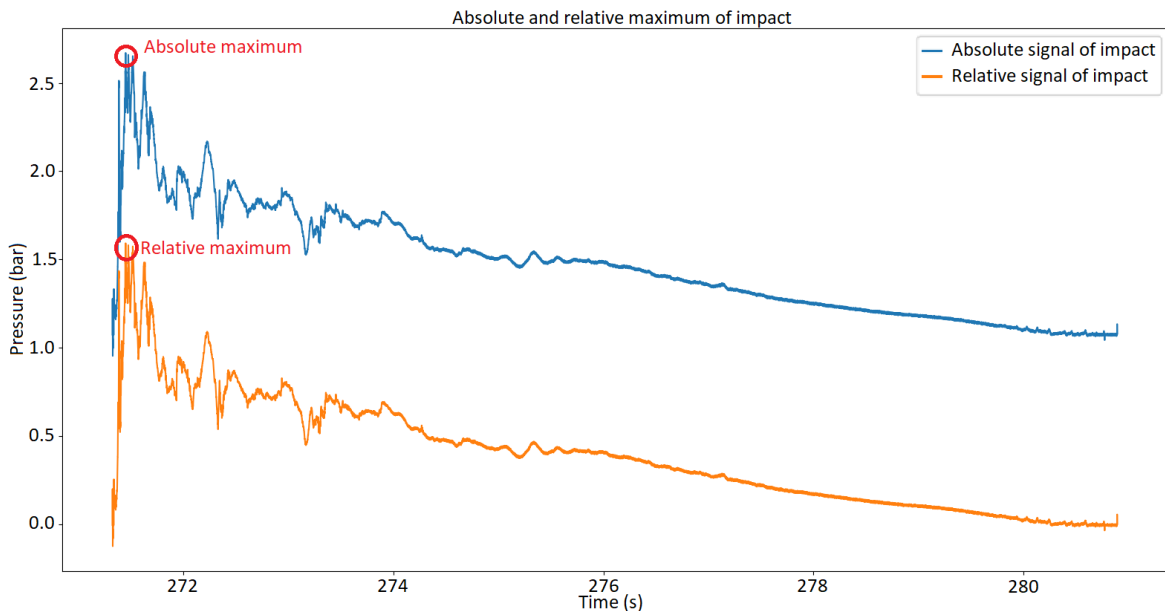


FIGURE 4.16 Maximum absolu et maximum relatif d'un impact.

4.3.3 Intégrale

A l'instar du temps de montée et du temps de descente, l'intégrale du signal d'impact est calculée entre 10% du maximum avant le maximum et 10% du maximum après le maximum. Pour ce calcul nous utilisons le signal relatif, c'est-à-dire le signal auquel nous avons retiré la composante continue. La figure 4.17 illustre ce critère.

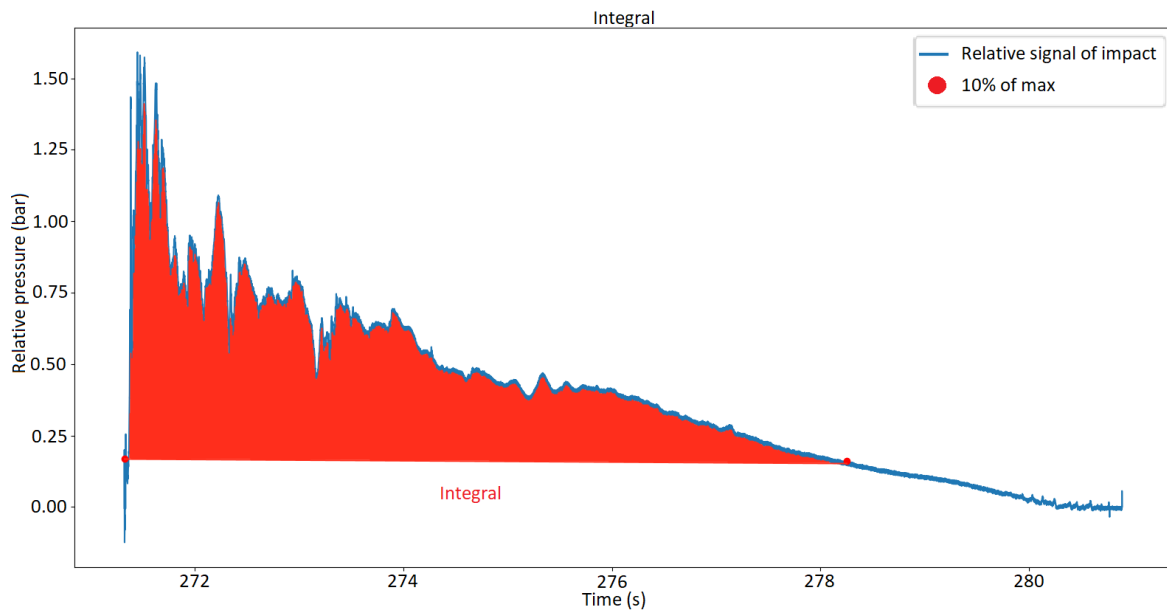


FIGURE 4.17 Intégrale de 10% du maximum avant le pic à 10% du maximum après le pic.

4.3.4 Impulsion de pression et durée d'impulsion

L'impulsion de pression correspond à l'intégrale de l'impact percussif. Il est donc possible de la calculer grâce à la formule suivante [38] [88] :

$$P(x, y) = \int_{ts}^{te} p(x, y, t) dt$$

où ts (*start*) correspond au début de l'impact percussif et te (*end*) à la fin de celui-ci. Dans la littérature, il a été déterminé que la durée d'impulsion est calculée grâce à la formule $2\Delta t$, où Δt correspond au temps de montée [27]. La figure 4.18 illustre l'impulsion de pression sur un impact.

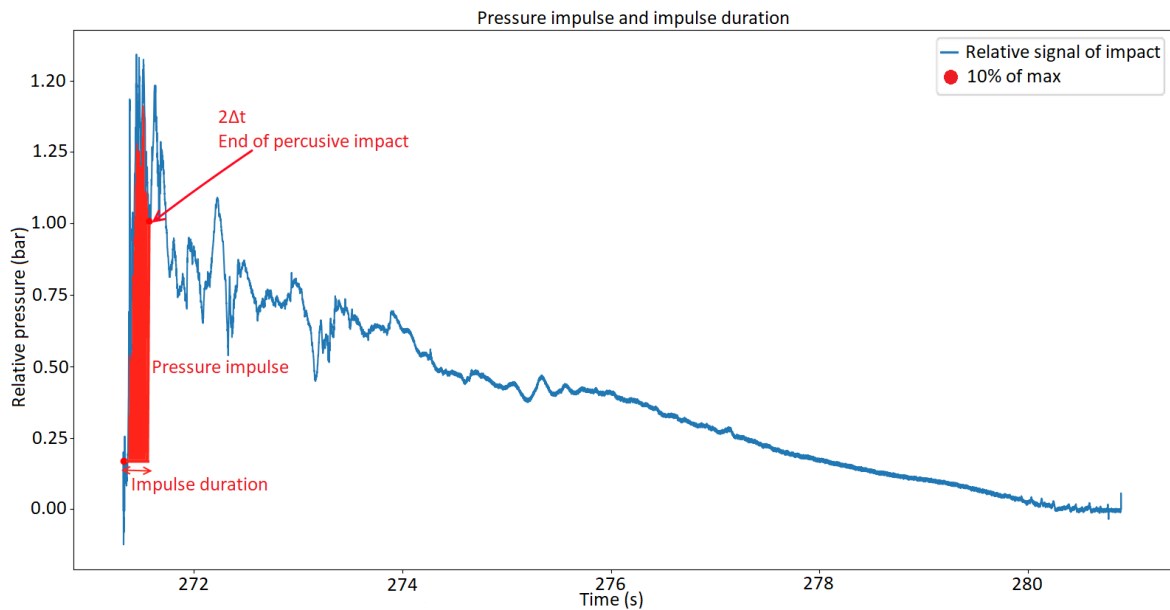


FIGURE 4.18 Impulsion de pression et durée d'impulsion.

Pour ce critère, nous avons été confrontés à deux problèmes majeurs. Le premier a été de déterminer si un impact contient un impact percussif et donc que l'impulsion de pression existe. En effet, pour certains impacts de vagues, il n'y a pas d'impulsion de pression. La figure 4.19 montre un impact pour lequel il n'y a pas d'impact percussif.

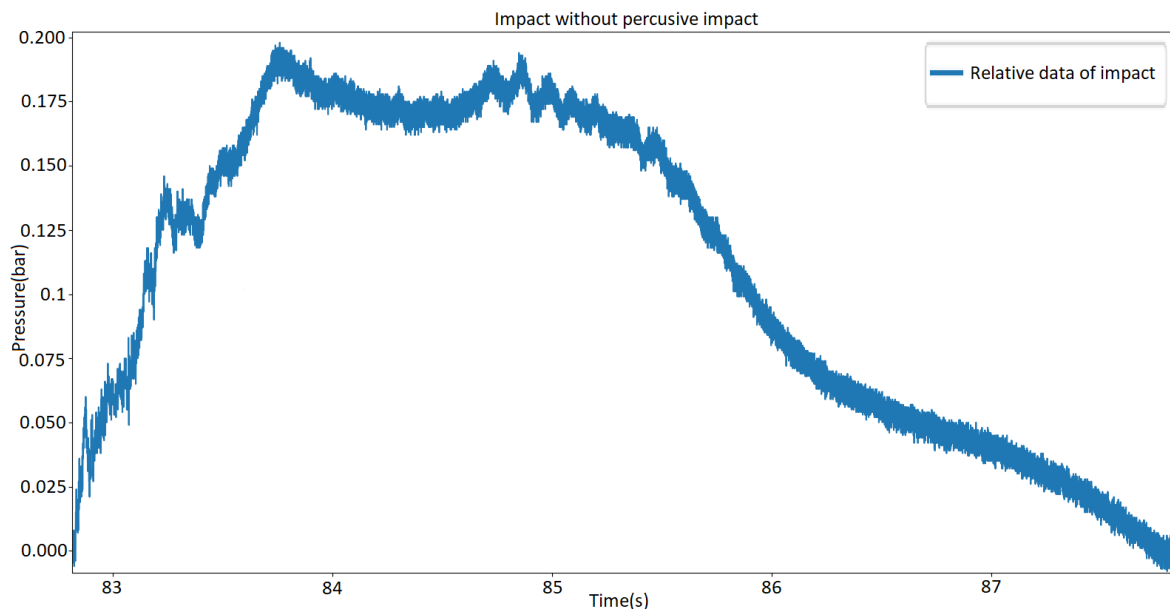


FIGURE 4.19 Impact sans impact percussif.

Nous avons essayé de déterminer de manière automatique si un impact était donc concerné par l'impulsion de pression ou non, mais malheureusement nous n'avons pas réussi à trouver de méthode fiable. Nous avons essayé de nous baser sur différents critères, comme le temps de montée ou encore la valeur maximale dans la dérivée, mais nous n'avons pas pu déterminer un seuil fixe suffisamment fiable. Nous avons de ce fait choisi de calculer l'impulsion de pression sur tous les impacts et laisser l'expert déterminer d'après son expertise si ce critère est pertinent ou non pour l'impact qu'il analyse.

Le deuxième questionnement que nous nous sommes posés concernant ce critère est comment déterminer la durée d'impulsion. Cela revient à connaître t_e . Après analyse des impacts enregistrés sur la digue de l'Artha, la formule de Bagnold n'est pas toujours juste et donne seulement une approximation. Nous avons de ce fait cherché une autre approche, notamment en nous basant sur la dérivée du signal. Cependant, nous n'avons, là aussi, pas trouvé d'approche fiable et avons donc utilisé la formule de Bagnold. La durée d'impulsion correspond donc ici à $2\Delta t$.

4.3.5 Force linéique maximale

Goda *et al.* [89] ont étudié la distribution de la pression d'impact. Ils font l'approximation d'un profil linéaire dans un contexte moyenné pour tester la stabilité de l'ouvrage. Dans notre cas, nous utilisons la distribution spatiale des capteurs de la croix pour estimer la force, en mètres linéaires, exercée par une vague sur l'axe vertical du mur. Ici, nous avons 11 capteurs qui sont placés sur un axe vertical. En prenant le capteur 0 (s_0) comme point d'origine, nous avons donc par capteur de l'axe vertical, les valeurs de hauteur suivantes :

- s_{10} : 3.65 m
- s_9 : 3.45 m
- s_8 : 3.25 m
- s_7 : 3.05 m
- s_6 : 2.85 m
- s_5 : 2.65 m
- s_4 : 2.45 m
- s_3 : 2.25 m
- s_2 : 2.05 m
- s_1 : 1.85 m
- s_0 : 0 m

Le calcul est réalisé en intégrant la distribution de pression verticale instantanée à l'aide de la méthode des trapèzes, supposant une linéarité de pression entre capteurs, grâce à la formule suivante :

$$F(x) = \int_{s_0}^{s_{10}} p(x) dx$$

La figure 4.20 montre la pression impact engendrée par une vague qui a touché les 11 capteurs situés sur l'axe vertical et la force associée (en noir sur la figure). L'axe des abscisses correspond au temps. L'axe des ordonnées de droite correspond à la pression en bar. L'axe des ordonnées de gauche correspond à la force linéique en kN/m.

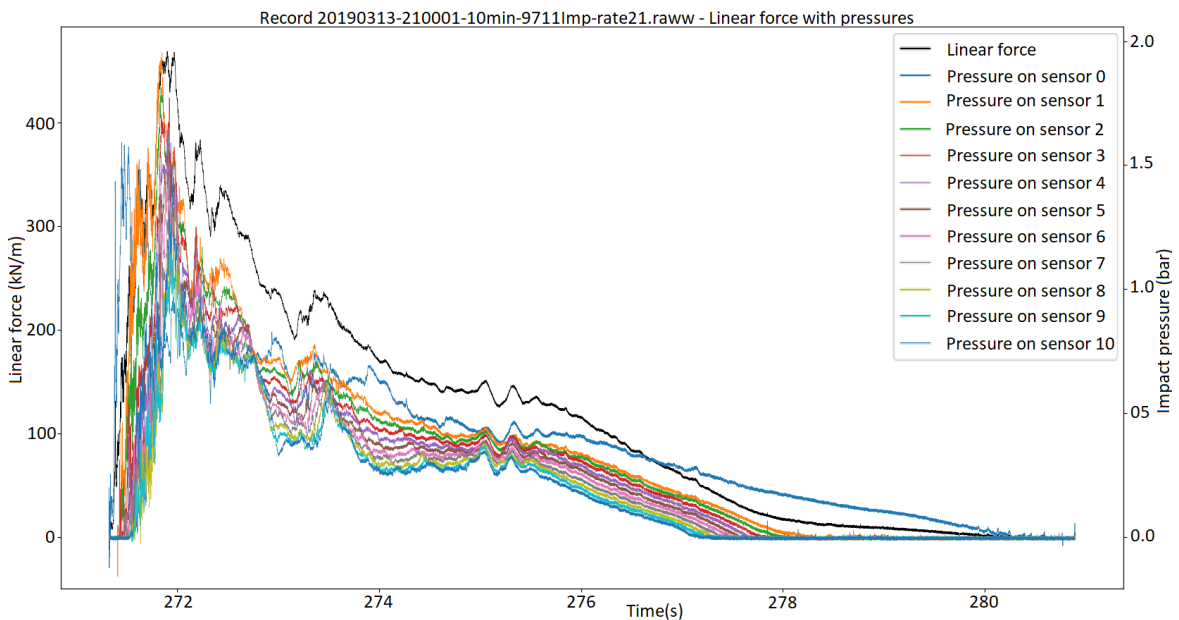


FIGURE 4.20 Force linéique et pression d'impact sur les 11 capteurs de l'axe vertical

Cette figure permet d'observer l'évolution de la force appliquée par l'impact de la vague sur la surface de la digue concernée. Cela est intéressant pour les experts car ça leur permet de mieux analyser la force appliquée lors d'un impact. Ce calcul de force linéique est effectué automatiquement pour tous les impacts détectés. L'objectif est de renseigner dans la base de données la force maximum engendrée par chaque impact. Ce calcul a en contrepartie un certain coût en termes de calcul, vu qu'il faut calculer la force linéique pour chaque instant t de l'impact. Un impact durant environs 10 secondes, avec $Fe = 10240$, cela fait un grand nombre d'intégrations à réaliser. Pour un enregistrement contenant des impacts de manière continue, soit environs 50 impacts en 10 minutes, il faut compter environ 40 secondes pour

le calcul de force sur tous les impacts. Dans l'absolu, ce temps est relativement rapide, mais par rapport au temps d'exécution de l'algorithme de détection des impacts et de calcul de critères, cela est relativement long.

4.3.6 Impacts associés

Bien souvent, les experts déterminent un impact à analyser en choisissant un critère spécifique. Cela conduit à sélectionner un impact et, pour l'analyser, ils sélectionnent le signal correspondant pour un capteur déterminé. Par exemple, ils vont choisir l'impact maximum enregistré par la station de mesure, qui a lieu sur le capteur 0 (capteur du bas) et analyser ce signal.

Cependant, il est très intéressant de pouvoir analyser en parallèle ce qu'il se produit sur les autres capteurs lors de cet impact. Jusqu'alors, l'expert devait ouvrir les signaux correspondants pour chaque capteur, et en extraire la partie qui correspond à l'impact qu'il a préalablement sélectionné. Sinon, il devait fouiller dans la base de données que nous avons réalisée afin de déterminer pour chaque capteur quel est l'impact correspondant et extraire dans la base de données la donnée pertinente (par exemple la pression maximum).

Afin de faciliter ce travail qui peut être fastidieux, nous avons voulu associer les impacts sur les différents capteurs qui correspondent à la même vague. La période moyenne des vagues étant généralement plus longue que la durée d'un impact, nous avons donc déterminé qu'en utilisant simplement l'inclusion et l'intersection temporelle entre les signaux des différents capteurs, nous serions capables d'associer les impacts correspondant à la même vague.

De ce fait, quatre cas de figure sont possibles :

- L'impact du capteur A est inclus dans l'impact du capteur B.
- L'impact du capteur A inclut l'impact du capteur B.
- Il existe une intersection entre la fin de l'impact du capteur A et le début de l'impact du capteur B.
- Il existe une intersection entre le début de l'impact du capteur A et la fin de l'impact du capteur B.

Ces quatre possibilités sont schématisées dans les figures 4.21 et 4.22 qui représentent respectivement l'inclusion et l'intersection.

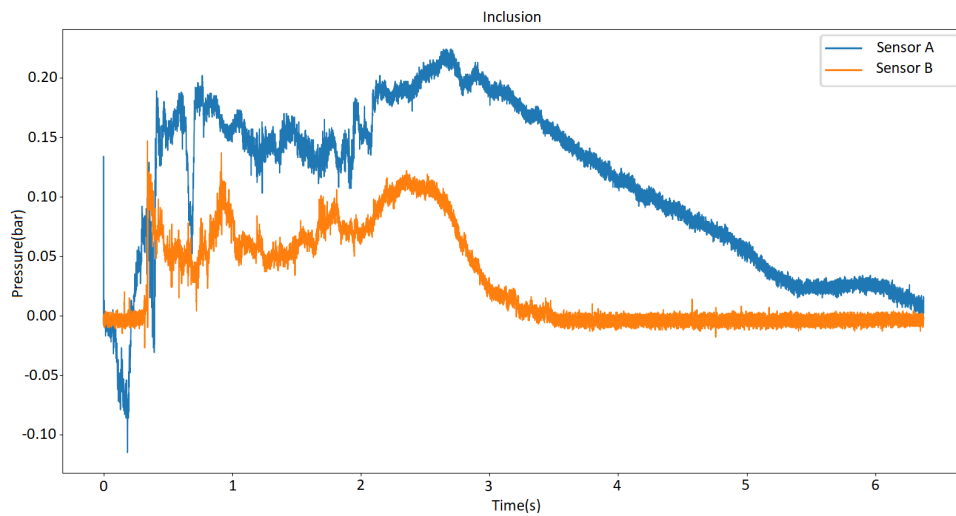


FIGURE 4.21 Inclusion temporelle d'un impact dans un autre.

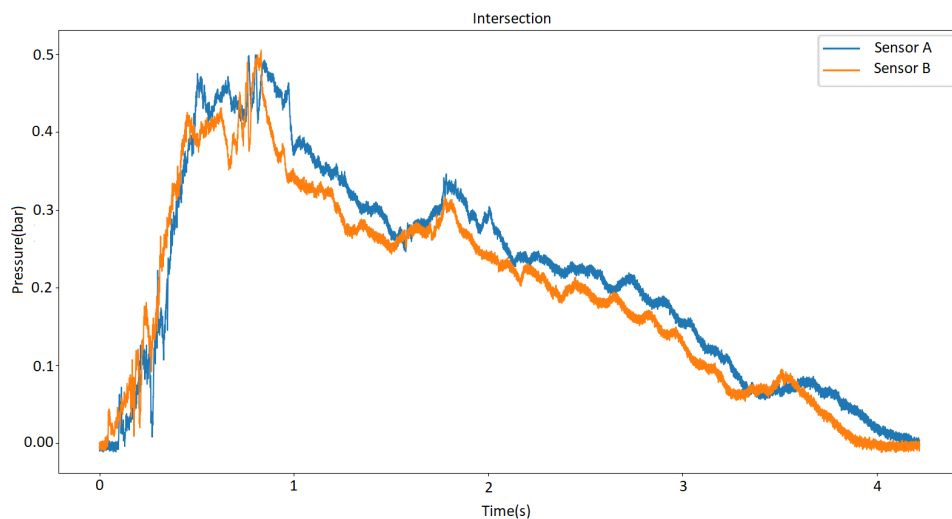


FIGURE 4.22 Intersection temporelle entre deux impacts.

Cela permet d'enregistrer directement dans la base de données, pour un impact sur un capteur, les identifiants des impacts qui sont liés à la même vague sur les autres capteurs. En effet, en utilisant un dictionnaire, il est simple d'enregistrer l'identifiant de l'impact associé par capteur sous le format *SensorId : UniqueId* où chaque combinaison est séparée par un point virgule, comme nous pouvons voir un exemple ci-dessous :

{1 : 1034;2 : 1043;3 : 1051;4 : 1057;5 : 1061;6 : 1065;11 : 1071;12 : 1073;13 : 1075;14 : 1078;15 : 1080}

Dans ce dictionnaire nous pouvons observer que l'impact que nous analysons est associé à l'impact 1034 qui a lieu sur le capteur 1, l'impact 1043 qui a lieu sur le capteur 2, et ainsi de suite. L'expert est donc capable de rapidement voir si l'impact de vague qu'il étudie a aussi impacté d'autres capteurs, et si oui, de retrouver rapidement les impacts concernés ainsi que leurs caractéristiques dans la base de données grâce à leur identifiant.

4.3.7 Bases de données

4.3.7.1 Base de données par impact

Tous ces critères qui sont calculés permettent de caractériser les impacts de vague. Cela permet aux experts de facilement extraire de l'information sur les impacts et donc de faciliter la manipulation des données enregistrées par la station de la digue de l'Artha.

Pour faciliter la visualisation et la recherche, nous souhaitions initialement utiliser une plateforme de Big Data. Une telle plateforme aurait eu un double intérêt. Le premier aurait été de pouvoir réaliser tous les traitements présentés auparavant de manière efficace. Le deuxième aurait été de pouvoir établir une base de données consultable en ligne, où il aurait également été possible de visualiser des signaux. Cependant, nous avons passé pas mal de temps sur la plateforme Big Data sélectionnée, mais nous n'avons pas eu les performances et fonctionnalités attendues. Tout d'abord, le tracé de courbes contenant beaucoup de points, comme c'est le cas pour nos signaux, n'était finalement pas réalisable. De plus, le temps de calcul n'était pas plus performant sur les serveurs de la plateforme choisie que sur l'ordinateur utilisé pour implémenter les algorithmes. Enfin, l'environnement de développement de cette plateforme a été à plusieurs reprises modifié entraînant des erreurs et des plantages sur des codes qui fonctionnaient lors de leur mise en production. De ce fait, par manque de temps, nous avons simplement choisi d'écrire les données dans un fichier type tableur.

Les figures A.6 et A.7 en annexe montrent un extrait de la base de données sous format tableur. Chaque ligne correspond à un impact et chaque colonne correspond à un critère. L'extrait sur la figure A.7 est la suite de l'extrait de la figure A.6.

Détails de chaque colonne :

- *FileName* : Nom du fichier de l'enregistrement,
- *SensorId* : Identifiant du capteur sur lequel l'impact s'est produit,
- *UniqueId* : Identifiant unique de l'impact,
- *StartIndex* : Indice de l'échantillon auquel l'impact débute (déterminé grâce à l'algorithme de la section 4.2.2.3),

- *StartTime* : Temps en seconde auquel l'impact débute dans l'enregistrement (*StartIndex/Fe*),
- *EndIndex* : Indice de l'échantillon auquel l'impact se termine (déterminé grâce à l'algorithme de la section 4.2.2.3),
- *EndTime* : Temps en seconde auquel l'impact se termine dans l'enregistrement (*EndIndex/Fe*),
- *MaxIndex* : Indice de l'échantillon où l'impact atteint son maximum,
- *MaxValue* : Valeur absolue du maximum de l'impact en bar (section 4.3.2.1),
- *RelativeMaxValue* : Valeur relative du maximum de l'impact en bar (section 4.3.2.1),
- *RiseTimeMax* : Temps de montée du signal d'impact entre le début et le maximum en seconde (section 4.3.1.1),
- *DropTimeMax* : Temps de descente du signal d'impact entre le maximum et la fin en seconde (section 4.3.1.2),
- *Integrale* : Intégrale du signal d'impact (section 4.3.3)
- *Duration* : Durée de l'impact en seconde (section 4.3.1.3),
- *PressureImpulse* : Impulse de pression (section 4.3.4),
- *ImpulseDuration* : Durée de l'impulsion en seconde (4.3.4),
- *StaticPressureBeforeImpact* : Pression statique (atmosphérique) avant l'impact en bar,
- *AssociatedImpacts* : Liste des impacts associés pour les autres capteurs (4.3.6),
- *AssociationImpactId* : Identifiant unique qui permet également de voir les impacts associés (tous les impacts ayant le même *AssociationImpactId* ont lieu lors de la même vague),
- *LinearForceMax* : Force linéique maximum de l'impact en kN/m (4.3.5),
- *IsError* : Permet de savoir si l'impact en question contient une erreur (pression atmosphérique qui varie lors de l'enregistrement de 10 minutes 4.2.2.3).

4.3.7.2 Base de données par enregistrement

Les critères présentés précédemment dans la section 4.3 permettent de caractériser les impacts qui sont enregistrés par la station de la digue de l'Artha. Mais ces critères permettent également d'obtenir de l'information à une échelle plus large. En effet, cela permet d'obtenir de l'information par enregistrement de 10 minutes. L'objectif est de réaliser une base de données par enregistrement et par capteur, permettant aux experts de sélectionner plus facilement et rapidement les enregistrements qui sont susceptibles d'être pertinents pour leurs études.

La figure A.8 en annexe montre un extrait de cette base de données pour le capteur 0. Le fichier Excel contient 16 onglets. Chaque onglet correspond à un capteur de pression et le nom de chaque onglet correspond à l'identifiant du capteur qu'il représente (0 à 15).

Détails de chaque colonne :

- *file* : Nom du fichier de l'enregistrement,
- *date* : Date à laquelle l'acquisition a été réalisée,
- *P1/10* : Moyenne des 10% plus grands pics de pression de l'enregistrement en bar,
- *Int1/10* : Moyenne des 10% plus grandes intégrales de l'enregistrement,
- *PresImp1/10* : Moyenne des 10% plus grandes impulsions de pression de l'enregistrement,
- *mean* : Moyenne du signal relatif en bar,
- *nbImpacts* : Nombre d'impacts détectés dans l'enregistrement,
- *Pmax* : Pression relative maximale en bar,
- *Tpmax* : Durée de l'impact où se produit *Pmax* en seconde,
- *Intpmax* : Intégrale de l'impact où se produit *Pmax*,
- *PresImppmax* : Impulsion de pression où se produit *Pmax*,
- *Patm* : Pression atmosphérique moyenne du signal en bar.

4.4 Calcul de covariance entre les capteurs

La covariance permet de quantifier l'écart entre deux signaux. La corrélation est une forme normalisée de la covariance. Ici, nous utilisons la covariance afin de déterminer l'écart de l'impact d'une vague sur deux capteurs différents. En effet, en calculant la covariance, nous pouvons connaître le décalage de l'impact entre deux capteurs en millisecondes. La figure 4.23 montre sur la colonne de gauche les signaux de deux impacts différents sur deux capteurs et sur la colonne de droite les covariances calculées pour ces signaux. Sur les figures de covariance, l'instant du maximum donne une estimation précise du délai de l'impact entre les deux capteurs [75]. Pour le premier impact, nous voyons qu'il y a un décalage de 282.5ms entre les deux capteurs. Pour le second impact, l'écart est de 138.9ms.

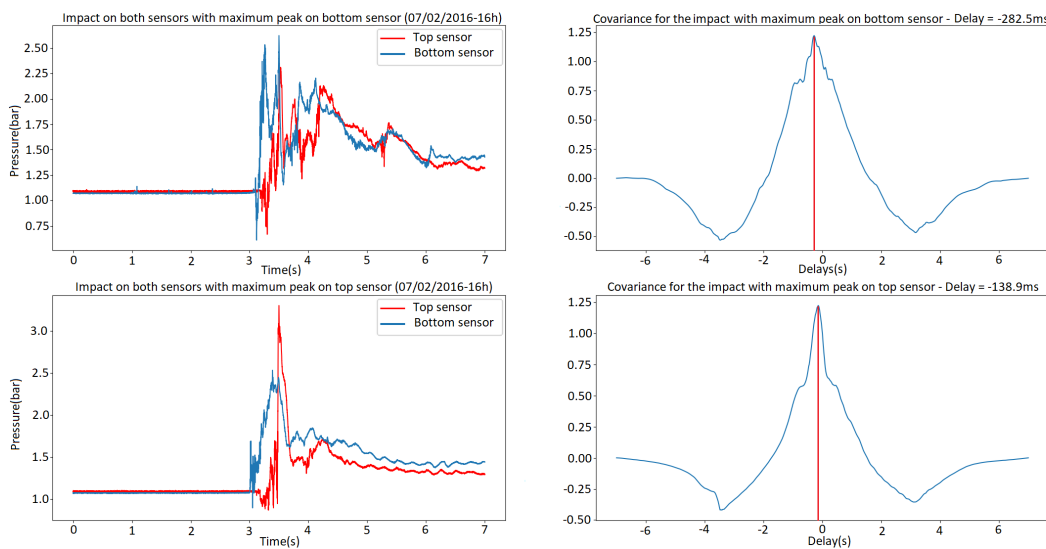


FIGURE 4.23 Covariance sur deux impacts.

Donc, en utilisant la covariance il est possible de connaître le décalage de l'impact d'une vague sur deux capteurs différents. Cela aurait pu être très pertinent pour notre étude, car ça nous aurait permis de calculer le décalage de l'impact de la vague entre tous les capteurs. Cependant, les capteurs sont placés trop hauts et de ce fait, la majorité des vagues impacts le capteur du bas, mais peu de vagues impactent les capteurs placés plus hauts. De ce fait, nous n'avons pas utilisé ces résultats pour la suite de nos travaux.

Il est également possible de calculer la covariance sur une partie de signal plus grande, contenant de ce fait plusieurs impacts, comme sur la figure 4.24.

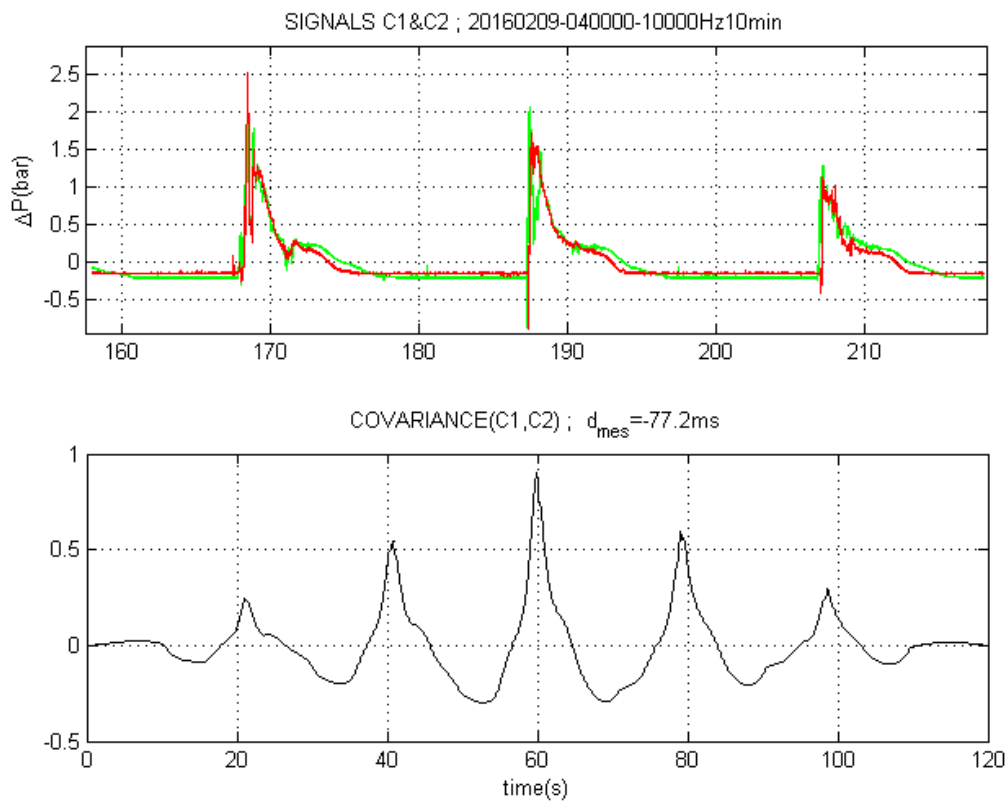


FIGURE 4.24 Covariance sur une partie de signal comprenant plusieurs impacts.

Cela permet par la suite de calculer la transformée de Fourier afin d'obtenir la densité spectrale de puissance des signaux, comme nous pouvons le voir sur la figure 4.25.

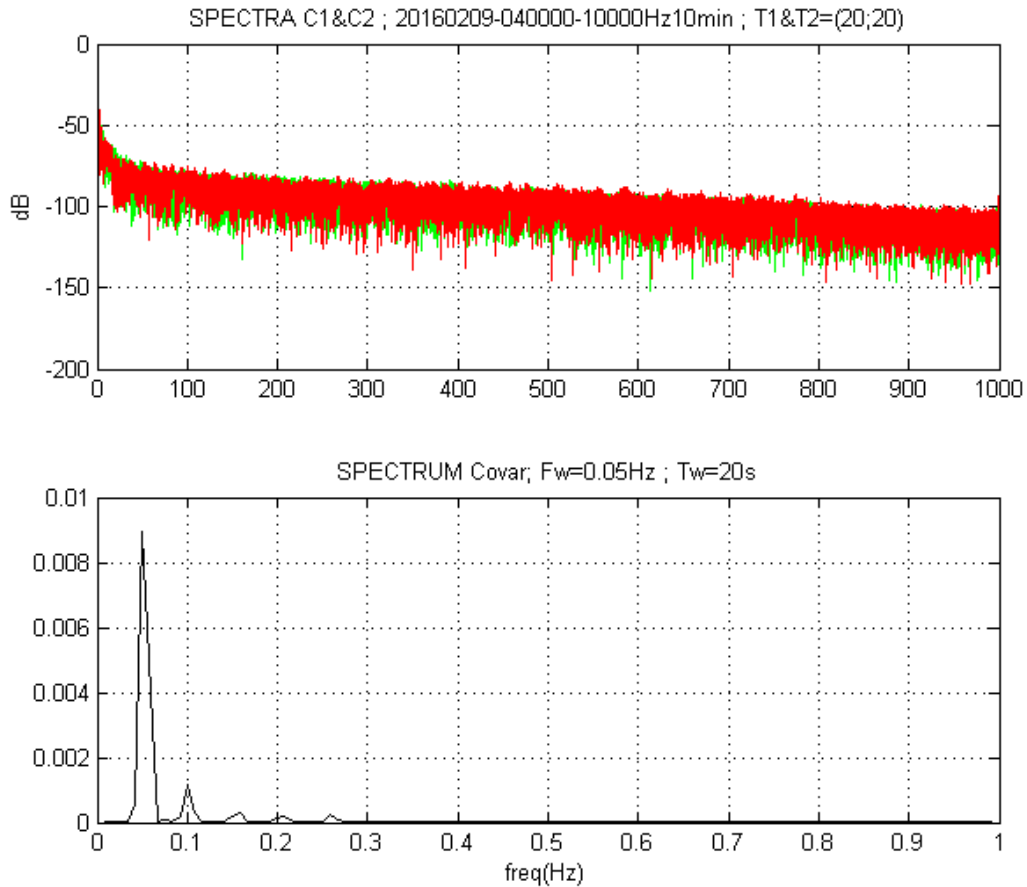


FIGURE 4.25 Spectre des signaux et spectre de la covariance.

Le pic principal exhibé sur le spectre de la covariance indique la fréquence d'impact des vagues. Ici la fréquence est $F_w = 0.05 \text{ Hz}$. Cela permet de connaître la période des vagues grâce à la formule suivante :

$$T_w = \frac{1}{F_w}$$

Ici, la période est donc de 20 secondes. Cette valeur peut être comparée et est en effet fortement corrélée avec la période des vagues mesurée par la bouée houlomotrice offshore.

4.5 Visualisation animée de la pression d'impact sur les capteurs

Il est assez simple de comprendre et d'analyser l'évolution de la pression d'un impact de vague sur un capteur en observant le signal correspondant. En effet, de cette manière nous pouvons bien examiner les changements et variations à chaque pas de temps. Cependant, faire la même chose sur plusieurs capteurs en parallèle afin de comprendre comment la vague a déferlé sur la digue est bien plus complexe. En effet, dû à la multitude de signaux superposés et à toutes les variations qu'ils contiennent, il devient difficile d'interpréter et de visualiser de quelle manière la vague a déferlé sur la digue. La figure 4.26 montre un impact de vague sur la digue de l'Artha qui a impacté les 16 capteurs de la station. Il est effectivement complexe, en observant cette figure, de visualiser comment la vague a déferlé sur le mur.

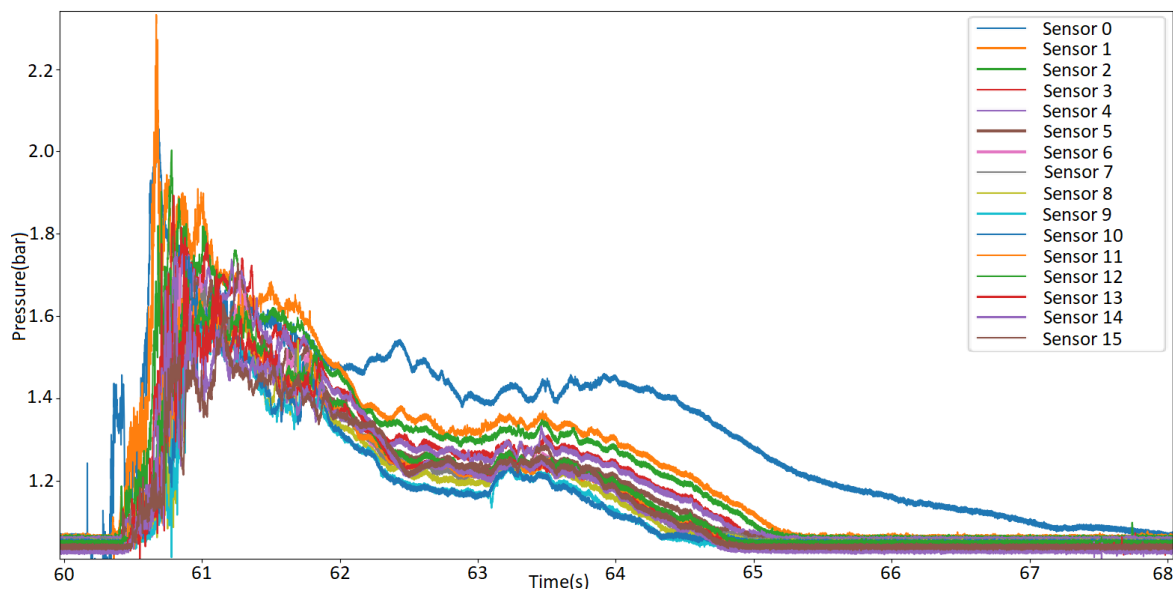


FIGURE 4.26 Impact d'une vague sur les 16 capteurs de la digue de l'Artha

Afin de mieux visualiser l'impact sur les différents capteurs, l'idéal serait d'avoir une visualisation vidéo dans le temps. C'est dans ce but que les deux sous parties suivantes ont été réalisées.

4.5.1 Evolution linéaire de la pression d'impact

L'objectif ici est de visualiser l'impact d'une vague sur les 11 capteurs placés sur l'axe vertical. Pour cela, il faut afficher pour chaque pas de temps, la valeur de la pression d'impact

de chaque capteur sur cet axe. La meilleure façon d'avoir un graphique visuel a été de placer en ordonnées l'identifiant des capteurs, et en abscisses la pression en bar. La figure 4.27 montre un tel graphique, où l'on peut observer la pression d'impact sur chaque capteur de l'axe vertical à 0.86 seconde après le début de l'impact.

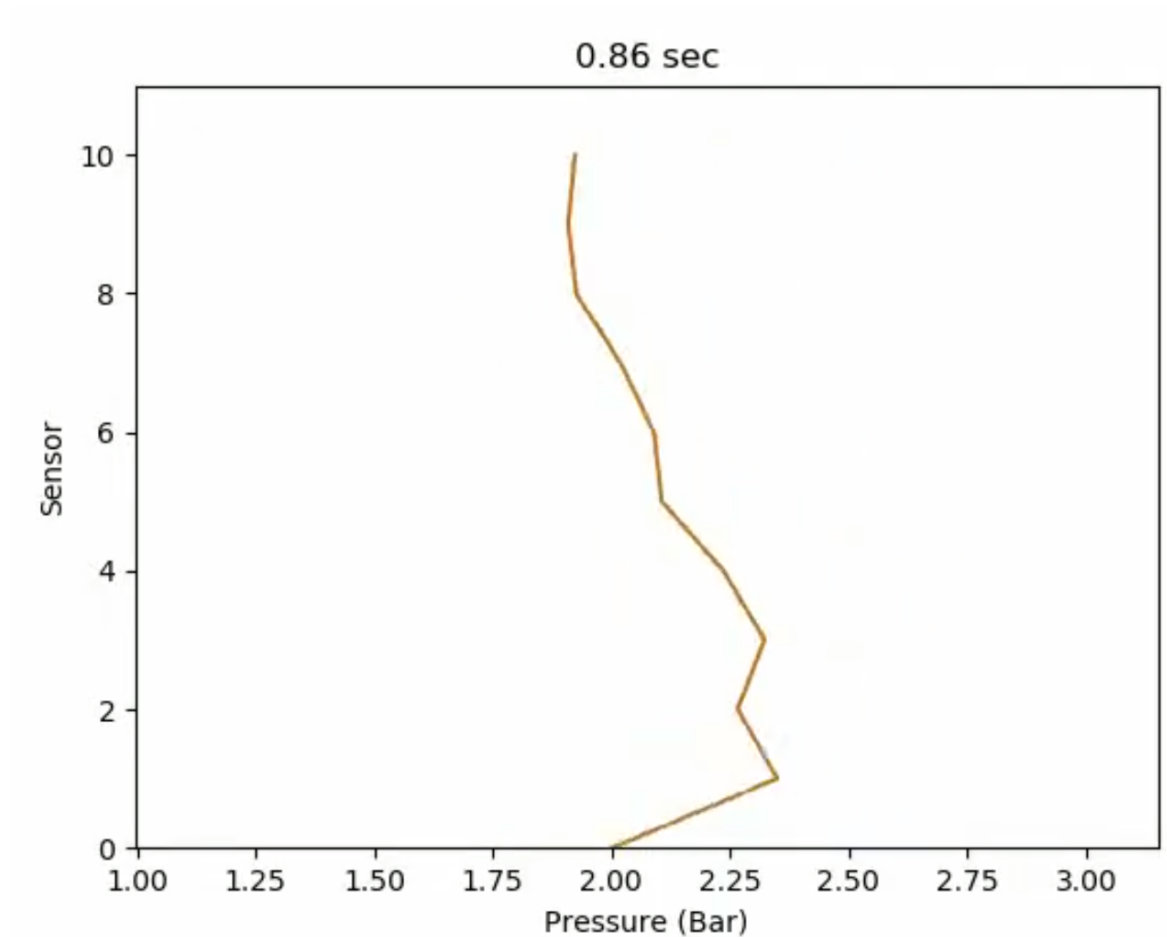


FIGURE 4.27 Capture de la vidéo à 0.86 seconde : graphique permettant la visualisation linéaire verticale de la pression sur les capteurs

Ce graphique est un extrait d'une vidéo qui permet de suivre l'évolution de l'impact dans le temps. En effet, afin de pouvoir visualiser l'impact dans son ensemble et non seulement à un instant t , il a été nécessaire de réaliser une vidéo qui permet d'observer l'évolution de la pression sur chaque capteur dans le temps. Un exemple de vidéo peut être visionné ici :

<https://www.youtube.com/watch?v=DW0UsGhOSj8>

Pour que cette vidéo soit utilisable, il a fallu l'enregistrer avec une fréquence d'images fixée à $Fps = 1000$ (Fps signifie *Frames per second*, ou image par seconde en français). En effet, nous avons la fréquence d'échantillonnage de la station $Fe = 10240Hz$. De plus, un impact dure en moyenne 10 secondes. De ce fait, il y a $10 * 10240 = 102400$ pas de temps. Pour observer l'évolution de l'impact dans le temps, il faut donc tracer autant de graphes. Il est donc nécessaire d'avoir Fps élevé, afin d'avoir une vidéo qui ne soit pas trop longue pour un impact de 10 secondes. Fps ne doit pas non plus être trop élevé afin que la personne qui analyse la vidéo soit capable d'observer les variations. Fixer Fps à 1000 nous a donc paru un bon compromis, permettant de voir l'impact à vitesse environ 1/10.

Le résultat permet en effet de mieux analyser et visualiser un impact sur l'axe vertical. Cependant, le problème est le temps nécessaire pour la création d'une telle vidéo. En effet, comme évoqué dans le paragraphe précédent, pour réaliser la vidéo il est nécessaire de créer un graphique pour chaque pas de temps. Pour cette vidéo nous avons donc tracé environ 100000 graphes que nous avons mis bout à bout pour réaliser la vidéo. Le temps de traitement pour cette vidéo a été d'environ 90 minutes. De ce fait, la réalisation automatique de vidéo pour tous les impacts est impossible. Par contre, il est envisageable de réaliser une vidéo pour un impact spécifique que les experts souhaiteraient analyser plus précisément.

4.5.2 Evolution 2D de la pression d'impact

L'objectif est de pouvoir visualiser la pression d'un impact sur les 16 capteurs en même temps. Cela signifie donc d'avoir une représentation 2D de l'impact, vu que les capteurs de la station de l'Artha sont positionnés en forme de croix. Nous ne pouvons donc pas utiliser la même approche que dans la partie précédente (4.5.1) qui permettait la visualisation 1D sur un axe vertical.

La meilleure solution a donc été de représenter graphiquement la croix où sont situés les capteurs et de faire varier la couleur des capteurs en fonction de la pression qui est appliquée dessus par l'impact de la vague. Pour compléter le graphique, afin que ce soit plus précis, nous avons choisi d'afficher la pression en bar à côté de chaque capteur. La figure 4.28 montre un exemple de graphique à 0.52 seconde après le début de l'impact. Le code couleur est divisé en 8, de vert clair à rouge foncé, permettant de visualiser l'intensité de la pression sur chaque capteur. Chaque couleur représente un intervalle qui est défini en fonction de la valeur maximum et minimum de l'impact.

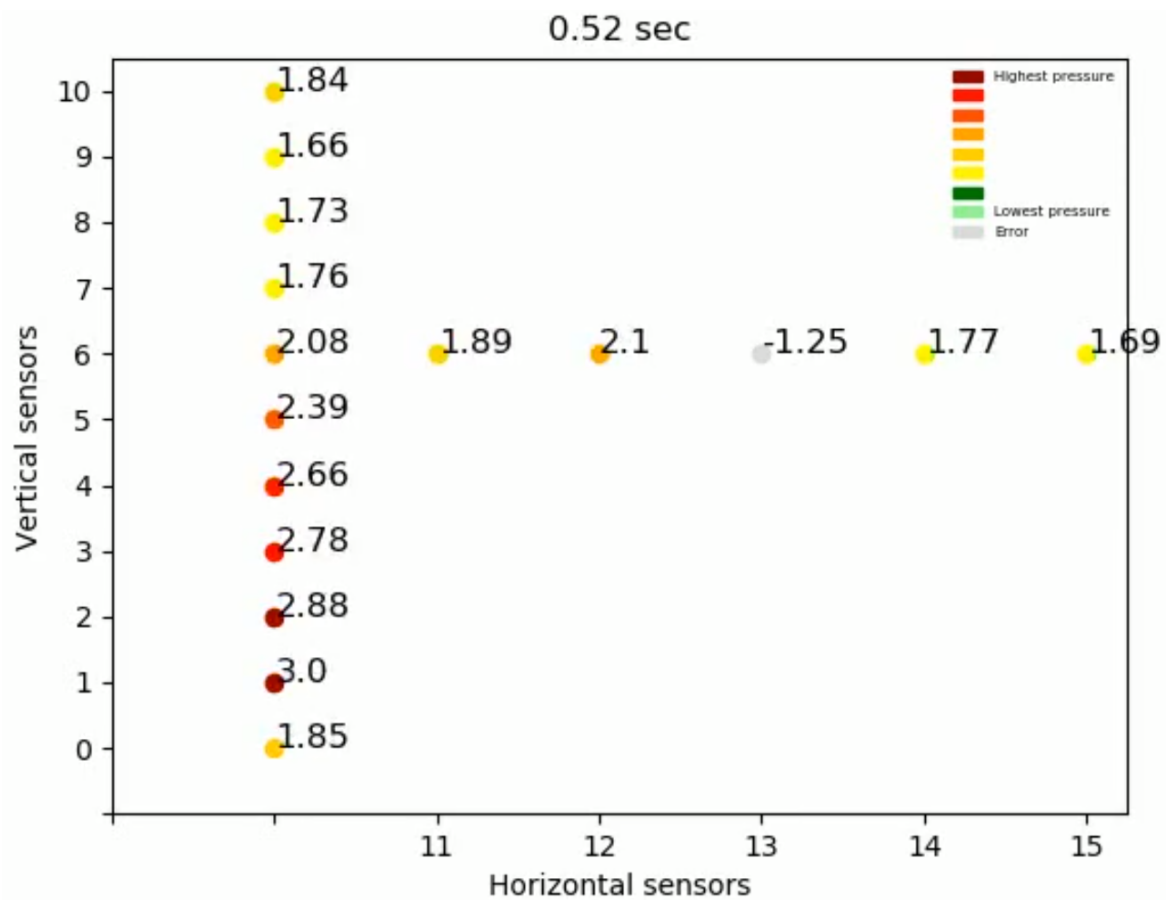


FIGURE 4.28 Capture de la vidéo à 0.52 seconde : graphique permettant la visualisation 2D de la pression sur les capteurs. La valeur à côté de chaque capteur correspond à la pression sur le capteur en bar. Le code couleur de vert à rouge permet d'avoir une indication de la force de la pression sur chaque capteur.

Ce graphique est un extrait d'une vidéo qui permet de suivre l'évolution de l'impact dans le temps. Afin de pouvoir visualiser l'impact dans son ensemble il a été nécessaire de réaliser une vidéo qui permet d'observer l'évolution de la pression sur chaque capteur dans le temps. Un exemple de vidéo peut être visionné ici :

<https://www.youtube.com/watch?v=9oqdKYHHCfc>

Tout comme pour la partie précédente (4.5.1), créer une telle vidéo prend beaucoup de temps et n'est donc pas faisable de manière automatique. Il est cependant possible de créer une vidéo pour un impact pertinent pour les experts, leur permettant de mieux visualiser l'impact de la vague sur la digue.

4.6 Retour sur la station de l'Artha

Cette section a pour objectif de justifier le choix fait dans les prochaines parties de n'utiliser que les données provenant du capteur 0.

Il est important de préciser que cette station n'a pas été positionnée idéalement. En effet, nous avons constaté que la croix contenant les capteurs est positionnée trop en hauteur par rapport au niveau d'eau moyen et donc seul le capteur du bas est réellement utilisé dans l'étude présentée ici. Pour détailler, l'histogramme de la figure 4.29 montre le nombre d'impacts total enregistrés par capteur entre juin 2018 et juin 2019.

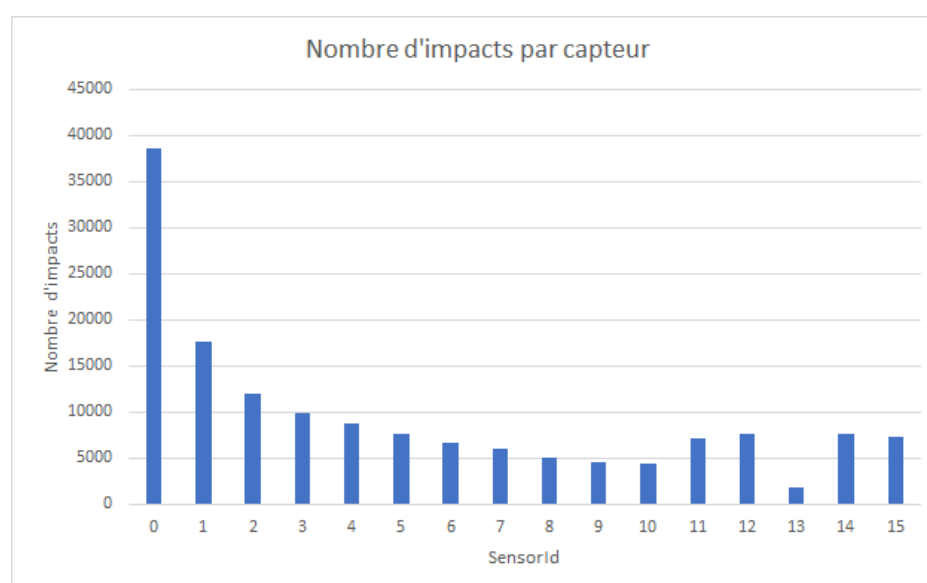


FIGURE 4.29 Histogramme montrant le nombre d'impacts par capteur entre juin 2018 et juin 2019.

L'histogramme de la figure 4.30 montre quant à lui le nombre d'impacts supérieurs à 0.3 bar en relatif enregistrés par capteur sur la même période.

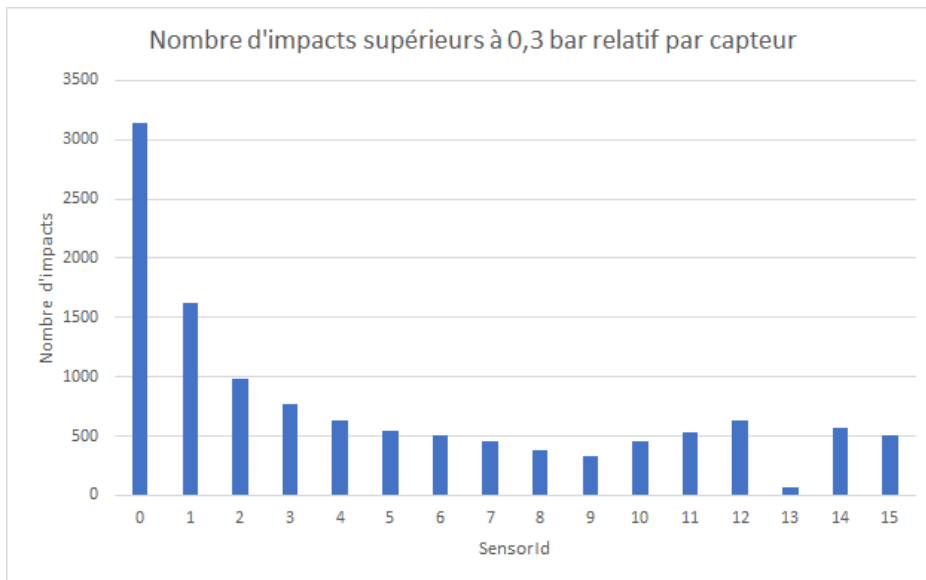


FIGURE 4.30 Histogramme montrant le nombre d'impacts supérieurs à 0.3 bar en relatif par capteur entre juin 2018 et juin 2019.

Ces deux graphiques montrent bien qu'il y a une grosse différence entre le nombre d'impacts enregistrés sur le capteur 0 et les autres, et amènent donc à se questionner sur le positionnement des capteurs. Cependant, il est nécessaire de préciser que l'emplacement de l'installation des capteurs a été dicté par le Conseil Départemental 64 (CD64) [82] qui est chargé des travaux sur les structures côtières, en invoquant l'intégrité de la structure. En effet, comme le montre la photo 4.31, seule la partie haute de la digue de l'Artha est verticale, la partie basse étant oblique. Afin de ne pas fragiliser la digue, les capteurs ont dû être installés sur la partie verticale.

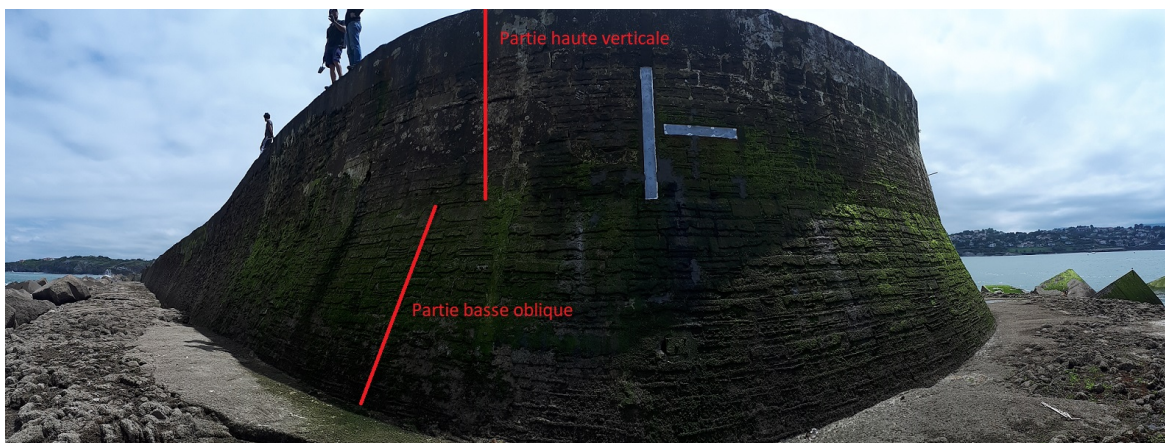


FIGURE 4.31 Photo du profil de la digue de l'Artha.

4.7 Prédiction de la pression d'impact des vagues

Cette partie a pour objectif d'extraire de l'information et de trouver des corrélations permettant de faire de la prédiction de pression d'impact des vagues sur la digue de l'Artha. Pour cela, nous allons suivre les étapes du processus KDD (section 2.2.1.3).

4.7.1 Etape 1 : Objectif

Dans l'optique de superviser la digue de l'Artha, nous souhaitons pouvoir prévoir la puissance d'impact des vagues sur la digue. En effet, si nous sommes capables d'anticiper les fortes pressions, alors il est possible de mieux superviser la structure.

L'objectif premier a donc été d'essayer de déterminer, pour les enregistrements de 10 minutes dont nous disposons, la valeur $P1/10$ (section 4.3.7.2) grâce aux données environnementales locales enregistrées au même moment. Il ne s'agit donc que d'une première expérience afin de valider si nous sommes capables, dans un premier temps, de déterminer la pression $P1/10$ dans un intervalle de 10 minutes grâce aux données environnementales enregistrées au même moment. Dans ce cas précis, il ne s'agit donc pas de la prédiction mais de vérifier s'il existe une corrélation entre la pression d'impact des vagues et les paramètres environnementaux.

4.7.2 Etape 2 : Les données

La donnée que nous cherchons à déterminer est la puissance $P1/10$ se produisant dans chaque enregistrement de la digue de l'Artha. Pour cela, nous avons besoin des données de la station de l'Artha. Nous avons donc utilisé la base de données par enregistrement présentée précédemment dans la section 4.3.7.2.

Afin de déterminer cette valeur $P1/10$ de chaque enregistrement, nous avons besoin des données environnementales enregistrées au même moment que ce dernier. Ces données proviennent des stations suivantes :

- La station météo de Socoa,
- Le marégraphe de Socoa,
- Le houlographe.

Ces données sont celles qui sont présentées dans la partie 3.2.3.

4.7.3 Etape 3 : Prétraitement des données

Pour cette étape, rien n'a été réalisé en ce qui concerne les données de la station de l'Artha, le prétraitement a été effectué avant l'algorithme permettant le découpage des impacts.

Pour ce qui est des données venant des autres stations, les prétraitements ont déjà été appliqués par les organismes possédant les données. Dans certains cas, comme pour les données du marégraphe (3.2.3.2), ce sera à nous de ne sélectionner que les données validées lors de l'étape suivante 4.7.4.

Il y a cependant une étape de prétraitement nécessaire pour les données provenant de la station météo de Socoa ainsi que du houlographe 06403. En effet, ces stations nous renseignent respectivement sur les valeurs *DD* et *thetap*. Ces valeurs nous donnent une direction (respectivement pour la direction du vent et la direction de la vague) sous deux formats différents :

- *DD* est une chaîne de caractères donnant la direction du vent. Il existe 16 possibilités pour cette chaîne de caractères. Par exemple, pour la direction nord cette chaîne de caractères est : "*Le vent qui souffle du nord*". Cette chaîne est du même type pour les autres directions. Les 16 directions possibles sont représentées sur la figure 4.32.

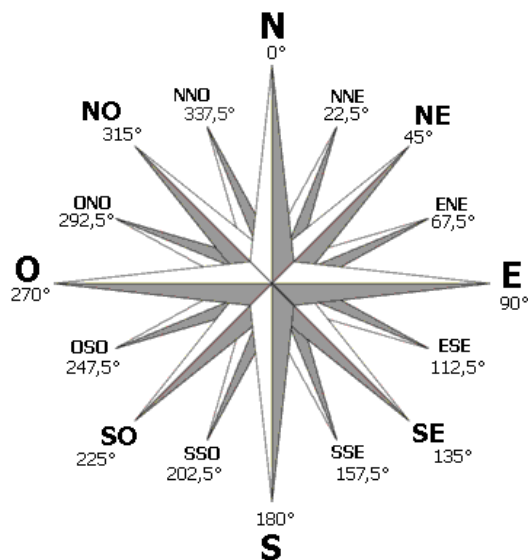


FIGURE 4.32 Rose des vents à 16 directions.

- θ est exprimée par un réel compris dans l'intervalle $[0, 360[$. Cela correspond à la direction de la vague en degrés par rapport au nord.

Il faut donc harmoniser ces deux champs pour qu'ils soient représentés de la même manière. De manière générale, les chaînes de caractères sont à éviter, il n'est donc pas envisageable de garder le format de DD . On pourrait donc penser à convertir les chaînes de caractères par leur équivalent en degrés. Cependant, pour un modèle d'apprentissage, le degré n'est pas une métrique intuitive. En effet, du fait du modulo 360 du degré d'une direction par rapport au nord, une valeur proche de 360 (exemple 359), exprime en réalité une direction très proche de 0. Or, en l'état, pour un système, 359 correspondra à un extrême, et 0 à un autre.

Nous avons donc décidé de représenter la direction du vent et des vagues grâce à des booléens. Chacun de ces paramètres est maintenant représenté grâce à 16 colonnes, ce qui correspond aux 16 directions existantes pour le vent. Chaque colonne prend comme valeur 0, sauf la colonne qui correspond à la direction effective du vent qui prendra la valeur 1. Pour les vagues, nous avons établi que chaque catégorie représente un intervalle de 22.5 degrés ($360/16 = 22.5$). De ce fait, nous avons pu définir un intervalle pour chacune des 16 directions : $]D - (22.5/2), D + (22.5/2)]$, où D correspond à la valeur de la direction en degrés. Exemple pour le nord, on a $D = 0$: $]0 - 11.25 \text{ mod } 360, 0 + 11.25 \text{ mod } 360]$, ce qui donne $]348.75, 11.25]$. Toutes les colonnes sont donc égales à 0, sauf la colonne qui représente l'intervalle dans lequel la valeur de θ est incluse qui prend la valeur 1.

4.7.4 Etape 4 : Format des données

Pour l'objectif de cette tâche, nous n'avons pas besoin de toutes les données présentées précédemment dans la section 4.7.2. Nous allons donc présenter ici les données que nous gardons pour notre étude.

Pour ce qui est des données de la digue de l'Artha, nous avons seulement besoin de la valeur $P1/10$ par enregistrement puisque c'est la valeur que nous cherchons à déterminer. Nous n'allons donc garder que cette valeur présentée dans la section 4.3.7.2. De plus, nous n'utiliserons que les données qui proviennent du capteur 0 (capteur du bas). Nous avons choisi le capteur 0 car c'est le seul qui est impacté par toutes les vagues qui percutent la digue. Les autres capteurs, étant plus hauts, ne sont impactés que par les plus grosses vagues.

Les données environnementales utilisées pour déterminer la pression maximum sont les suivantes :

- Pour les vagues :

1. h_{110d} ,
 2. th_{110d} ,
 3. $thetap$ (représenté par les 16 booléens décrit dans la partie 4.7.3).
- Pour le vent :
1. DD (représenté par les 16 booléens décrit dans la partie 4.7.3),
 2. Ff .
- Pour le marégraphe :
1. *Valeur* du niveau d'eau.

Les figures B.1 et B.2 en annexe montrent un extrait de données formatées.

4.7.5 Choix de l'approche

Larroque *et al.* [74] ont calculé un modèle linéaire multivarié en utilisant les données environnementales de vent, de vagues et de niveau d'eau. L'objectif était de déterminer les combinaisons des paramètres environnementaux engendrant des conditions favorables aux plus grandes pressions d'impact. Cette étude a permis de déterminer que ces variables expliquent 58% de la variabilité des données. Un test ANOVA de ce modèle multivarié a montré, comme prévu, que la variable la plus importante est la hauteur des vagues expliquant 45% de la variabilité totale. Le niveau d'eau explique quant à lui 8,5%, la direction de la houle seulement 3,7% et la période moins de 0,5% de la variabilité totale. L'influence du vent ne semble pas significative dans ce modèle. La figure 4.33 montre un graphique qui synthétise ces résultats.

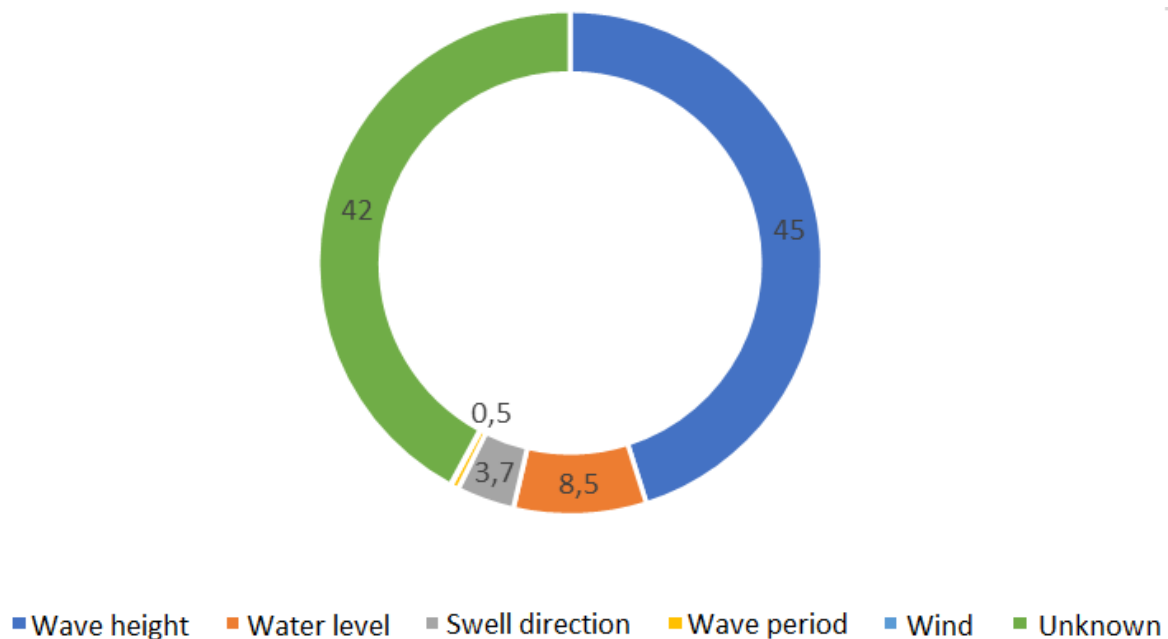


FIGURE 4.33 Graphique représentant l'incidence de chaque variable environnementale sur la pression d'impact des vagues (en pourcentage).

Nous avons donc choisi une approche différente. Le *Machine Learning* (ML), le *Deep Learning* et plus globalement l'Intelligence Artificielle sont des méthodes d'apprentissage qui permettent d'identifier des corrélations délicates dans des ensembles de données pour lesquels l'analyse exploratoire n'a pas permis de déterminer correctement la forme du modèle sous-jacent. L'objectif n'est pas de donner une formule explicite pour la distribution des données, mais plutôt que l'algorithme détermine le modèle par lui-même, directement à partir des données. Le but du ML n'est donc pas d'arriver à une connaissance des données mais plutôt à un modèle exploitable et reproductible. De plus, le ML est inhérent à l'informatique et à l'optimisation, qui permet un apprentissage rapide sur de gros jeux de données.

4.7.6 Choix de l'algorithme

Nous avons choisi d'utiliser les réseaux de neurones artificiels. Ces réseaux ont la capacité d'apprendre par eux-mêmes. En effet, ils sont capables d'apprendre grâce à un ensemble d'exemples et de s'autocorriger lorsqu'ils se trompent. Ils sont souvent utilisés pour résoudre des problématiques complexes de manière efficace.

Les réseaux de neurones sont très bons pour les exercices de prévisions, ce que nous souhaitons réaliser à terme si ce premier modèle fonctionne.

De plus, ces réseaux sont capables de travailler sur des événements en temps réel. De ce fait, dans le cas où nous arrivons à établir un modèle qui fonctionne correctement, cela nous permettrait de l'utiliser en temps réel.

Pour finir, les réseaux de neurones artificiels sont capables de travailler avec des connaissances incomplètes et ont une tolérance à l'erreur des entrées, ce qui les rend très intéressants lorsque certaines données sont manquantes ou erronées.

4.7.7 Etape 7 : Application de l'algorithme

Nous avons mis en place un réseau de neurones artificiels à propagation avant, *feed-forward neural network* en anglais. Lorsque l'on configure un réseau de neurones, il y a certains paramètres que nous pouvons déterminer assez rapidement en fonction du besoin. Par exemple, le nombre de neurones en entrée dépend directement des données, la fonction de coût ne sera pas la même si nous faisons de la classification ou si nous faisons de la régression. Mais il y a beaucoup d'autres paramètres qui sont plus compliqués à fixer, comme le nombre de couches cachées, le nombre de neurones par couche, les fonctions d'activation des neurones, etc. Pour ces paramètres, il n'y a pas d'autres choix que de faire différents tests, afin d'optimiser au mieux le modèle.

De ce fait, nous avons lancé de nombreux tests et nous avons conclu que la meilleure configuration pour notre réseau de neurones était la suivante :

- Couche d'entrée : 36 neurones représentant les 36 *features* des données présentées dans la partie 4.7.4,
- 4 couches cachées :
 - Couche 1 : 85 neurones cachés,
 - Couche 2 : 85 neurones cachés,
 - Couche 3 : 65 neurones cachés,
 - Couche 4 : 44 neurones cachés.
- Fonction d'activation des neurones des couches cachées : redresseur,
- Couche de sortie : 1 neurone représentant la valeur $P1/10$ déterminée par le modèle,
- Fonction d'activation de la couche de sortie : linéaire,
- 10% de chance de désactivation pour chaque neurone, afin d'éviter le surapprentissage lors de la phase d'entraînement,
- Fonction de coût : *Root mean squared error* ($RMSE = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{n}}$),
- Algorithme d'optimisation : *Adam*, un algorithme reconnu pour son efficacité, et qui a donné les meilleurs résultats lors de nos tests,

- Taille des lots : 1, permettant un meilleur renforcement car la *back propagation* est réalisée après chaque exemple,
- Nombre d'époques : 150, permettant de bien ajuster le modèle.

4.7.8 Etape 8 : Interprétation des résultats

Lors de l'entraînement, nous avons réalisé une validation croisée avec $k = 10$, donc 10 itérations. Cela a permis de calculer une variance moyenne du modèle de 0.078 avec un écart type de 0.015. Nous pouvons donc en déduire que l'apprentissage du modèle semble se faire correctement et de manière robuste.

Lors de la validation du modèle grâce au jeu de test, nous obtenons une *median absolute error* ($MedAE(y, \hat{y}) = \text{median}(|y_1 - \hat{y}_1|, \dots, |y_n - \hat{y}_n|)$) très faible (0.0008) et une *mean absolute error* ($MAE(y, \hat{y}) = \frac{\sum_{i=1}^n (|y_i - \hat{y}_i|)}{n}$) également faible (0.07), ce qui suggère que le modèle fait de bonnes prédictions. Cependant, nous avons un *mean squared error* ($MSE(y, \hat{y}) = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}$) de 0.02, ce qui semble faible mais lorsque nous calculons sa *root mean squared error* ($RMSE(y, \hat{y}) = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}}$) nous obtenons 0.14. Ceci montre qu'en réalité le modèle fait de grosses erreurs. En effet, si nous observons l'histogramme de la figure 4.34, nous pouvons observer que le modèle fait beaucoup de bonnes prédictions sur le jeu de test, mais que les erreurs qu'il commet sont grandes (à l'échelle du phénomène étudié).

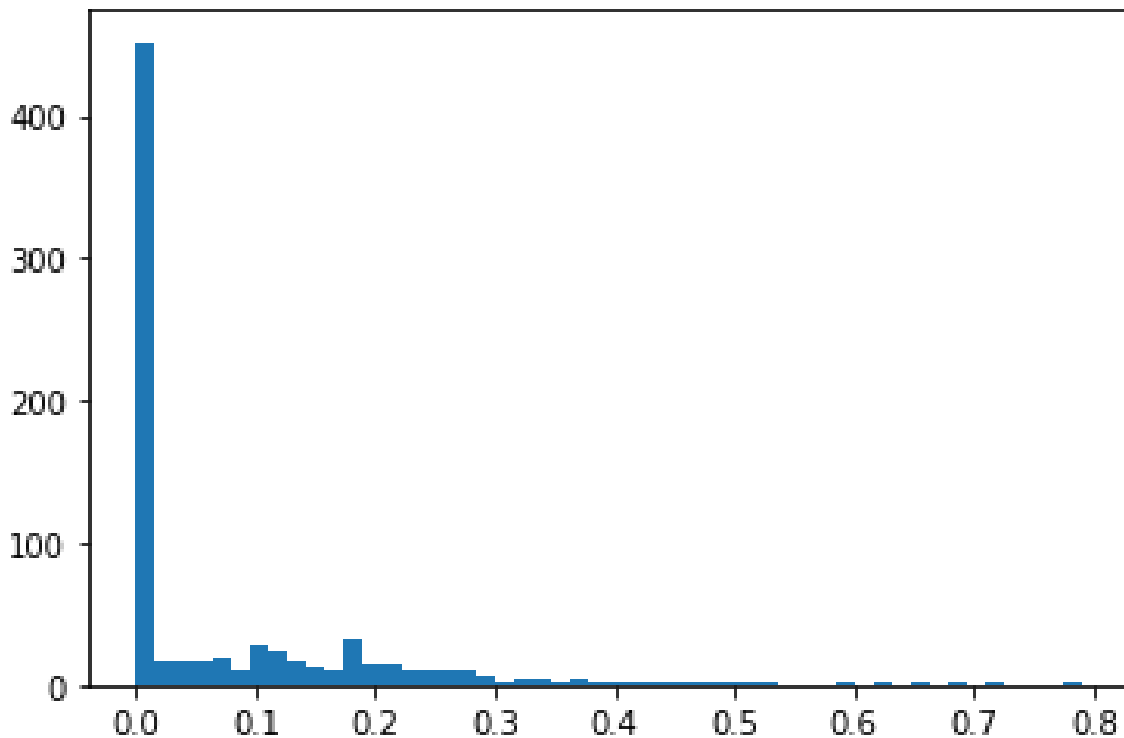


FIGURE 4.34 Histogramme représentant les erreurs du modèle par intervalles. Valeur des erreurs en abscisses et nombre d'exemples en ordonnées.

Ceci est vérifiable par l'analyse des résultats obtenus lors de la validation du modèle avec le jeu de validation (Annexe B). De plus, lors de l'analyse de ces résultats, nous pouvons observer que la majorité des bonnes prédictions sont faites pour des valeurs proches ou égales à 0. Or dans notre cas, nous souhaitons principalement être capables de déterminer les grandes valeurs, qui correspondent à de grosses pressions d'impacts sur la digue.

De ce fait, ce modèle n'est pas valide puisqu'il fait de grosses erreurs. Ceci peut être dû au volume important d'exemples dont la valeur $P1/10$ est égale à 0, mais il s'agit de données collectées sur le terrain, et donc qui représentent la réalité que le modèle devrait être capable de déterminer.

4.7.9 Conclusion

Ce modèle avait pour objectif de tester si nous étions capables de déterminer la pression d'impact des vagues sur la digue de l'Artha grâce aux données environnementales. Il s'agissait d'une première étape qui, en cas de succès, aurait permis de pousser plus loin afin d'essayer de faire de la prévision de pression d'impact.

Nous devons malheureusement conclure des résultats obtenus que le modèle que nous avons entraîné ne permet pas à déterminer la pression d'impact des vagues grâce aux paramètres environnementaux locaux. En effet, ce modèle commet de trop grosses erreurs, qui le rend inutilisable. Cela rejoint les résultats obtenus par Larroque *et al.* [74], et que donc d'autres paramètres entrent en compte pour déterminer la puissance d'impact des vagues.

Nous avons donc cherché à comprendre ce qui pouvait avoir un rôle sur la variation des pressions d'impacts. Cette recherche a donné lieu à la section suivante 4.8.

4.8 Clustering des types d'impacts des vagues

Cette partie a pour objectif d'extraire de l'information permettant de regrouper les impacts afin d'établir différents clusters. Pour cela, nous allons ici aussi suivre les étapes du processus KDD (section 2.2.1.3).

4.8.1 Etape 1 : Objectif

Comme nous l'avons vu dans la section précédente 4.7, il est complexe de déterminer la pression d'impact d'une vague en se basant seulement sur les données environnementales. Nous avons donc cherché à comprendre quels autres facteurs pouvaient engendrer cette variabilité, et nous avons fini par étudier les types d'impact des vagues. En effet, en fonction de la manière dont une vague déferle sur une structure, la pression appliquée par l'impact sera plus ou moins forte (section 2.1).

Nous avons validé cette hypothèse en observant certaines pressions d'impact puissantes qui ont eu lieu lors de conditions météorologiques relativement calmes. En effet, nous pouvons montrer comme exemple la pression d'impact la plus grande que nous avons enregistrée depuis la mise en marche de la station de mesure. Une pression de 5.512 bars (valeur absolue, ce qui équivaut à 4.469 bars en valeur relative) a été enregistrée sur le capteur 0 lors de l'acquisition du 21/01/2019 entre 4h et 4h10 du matin. A ce moment-là, la plus grande vague enregistrée par le houlographe avait une hauteur de 2.29m, donc une "petite" vague. En parallèle, la station météo de Socoa enregistrerait un vent faible, de seulement 3 mètres par seconde de moyenne. De plus, toutes les autres vagues ont appliqué de faibles pressions sur la digue durant les 10 minutes d'acquisition. La figure 4.35 montre l'acquisition dans son ensemble, et la figure 4.36 zoom sur l'impact le plus puissant.

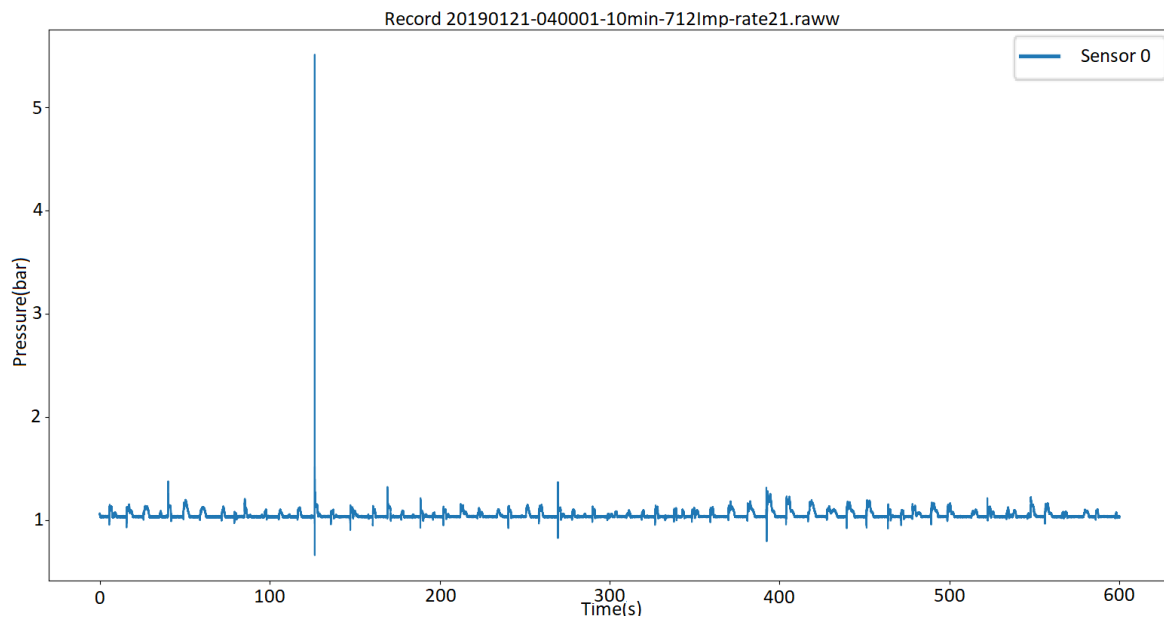


FIGURE 4.35 Enregistrement de 10 minutes contenant la pression d'impact la plus puissante.

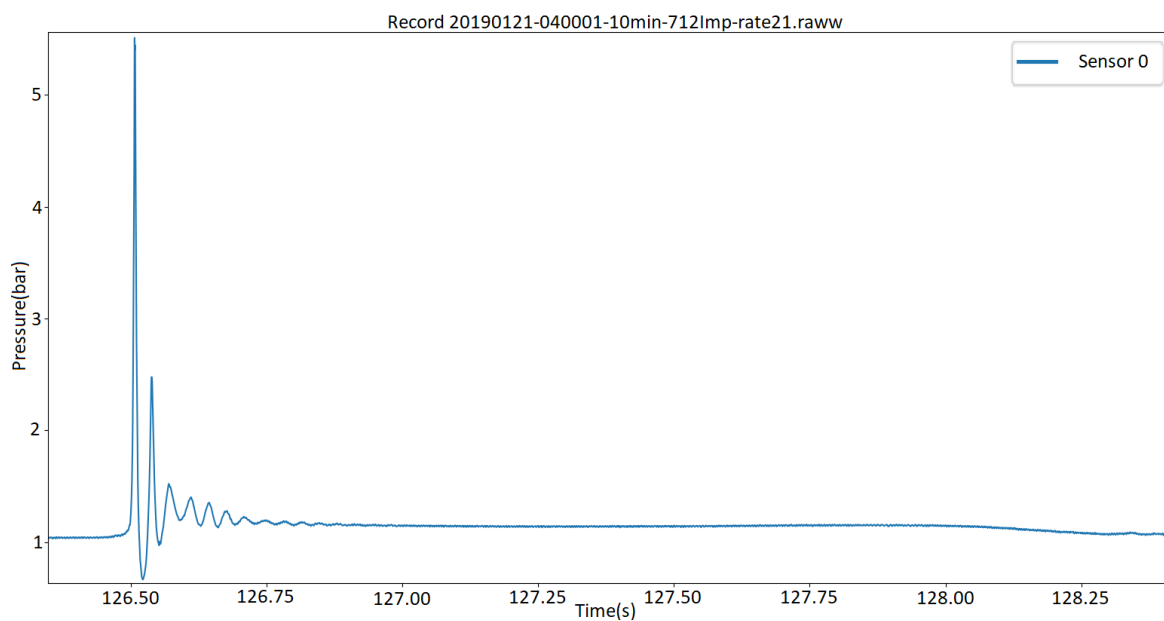


FIGURE 4.36 Pression d'impact la plus puissante enregistrée.

Il y a donc un intérêt à étudier les types d'impact des vagues qui déferlent sur la digue de l'Artha. L'objectif ici est d'essayer de regrouper ces impacts par classes, comme nous l'avons vu dans la section 2.1. Cependant, plusieurs problématiques se posent.

En effet, les études expérimentales en laboratoire sont souvent en 2D alors que le processus naturel est entièrement en 3D, ce qui pourrait conduire à une sur ou sous-estimation de la pression maximale [36]. L'échelle est différente entre le phénomène réel et les expériences en laboratoire même si les expériences récentes à grande échelle se rapprochent de l'échelle réelle [32]. Une mise à l'échelle est donc nécessaire, mais comme l'eau de mer est difficile à utiliser dans les expériences en laboratoire, un fluide différent est utilisé. La différence entre les fluides impliqués rend difficile l'établissement d'une règle de mise à l'échelle [31]. De même, les vraies structures côtières sont souvent protégées par une barrière de blocs en béton qui diminue la force des vagues. Enfin, les variables environnementales telles que la magnitude et la direction du vent, le niveau d'eau et la direction des vagues ne sont généralement pas prises en compte dans les études en laboratoire.

De plus, les études sur le terrain n'ont jamais été menées sur de longues périodes. Or cela est nécessaire afin de pouvoir enregistrer des données avec diverses conditions météorologiques et diverses combinaisons de paramètres environnementaux.

Pour ces raisons, la classification existante établie en laboratoire ne convient pas. De ce fait, nous ne sommes pas capables de classifier les impacts qui se produisent sur la digue de l'Artha. C'est pourquoi nous souhaitons réaliser un regroupement (*clustering*) de nos impacts sur la digue de l'Artha. Nous avons obtenu des résultats préliminaires [90] en n'utilisant que les plus gros impacts, ce qui ne concerne finalement que peu d'impacts. Cependant, il y a intérêt à établir des clusters sur un ensemble de données plus larges, c'est ce que nous souhaitons faire ici.

4.8.2 Etape 2 : Les données

Les données utilisées pour ce travail sont les données des impacts de la digue de l'Artha. Ces données ont été présentées dans la partie 4.3.7.1.

4.8.3 Etape 3 : Prétraitement des données

Ici, aucun prétraitement additionnel n'est réalisé. En effet, nous travaillons avec les données d'impacts qui proviennent des signaux que nous avons déjà prétraités.

En revanche, nous n'avons gardé que les données pertinentes pour notre recherche. Premièrement, nous avons sélectionné seulement les impacts qui se sont produits sur le capteur 0. Ce capteur est le plus bas et est donc celui qui est impacté par toutes les vagues. Nous aurions pu essayer de faire du clustering des impacts en prenant en compte tous les capteurs. Le problème est que la plupart des impacts ne touchent pas tous les capteurs, ce qui

rendrait la tâche complexe et donnerait probablement des résultats incohérents. De ce fait, nous avons préféré nous concentrer seulement sur les impacts du capteur 0, ce qui représente 38671 impacts pour la campagne 2018-2019.

Ensuite, nous avons choisi de supprimer tous les impacts dont la colonne *IsError* est à "VRAI". Cela permet de supprimer quelques erreurs / valeurs aberrantes. Après ce tri, il reste 38634 impacts.

Pour finir, nous avons choisi de supprimer les impacts trop faibles qui sont souvent très chaotiques et donc qui pourraient altérer les regroupements. De ce fait, nous n'avons gardé que les impacts dont le pic est supérieur à 0.3 bars en relatif. Après ce dernier filtre, la majorité des impacts étant très faibles, il ne reste que 3112 impacts.

4.8.4 Etape 4 : Format des données

Comme précisé dans la section 4.8.2, nous utilisons les données des impacts que nous avons extraites des acquisitions de la digue de l'Artha. En revanche, nous ne gardons pas tous les *features* car certaines sont inutiles pour notre travail actuel. Nous avons donc gardé les colonnes suivantes :

- RelativeMaxValue,
- RiseTimeMax,
- DropTimeMax,
- Integrale,
- Duration,
- PressureImpulse,
- ImpulseDuration.

4.8.5 Etape 5 : Choix de l'approche

Comme nous l'avons expliqué dans la section 4.8.1, il est impossible de classer nos impacts en utilisant la classification existante. De ce fait nous devons créer nos propres *clusters*.

Le premier problème est que nous avons un nombre important d'impacts et donc réaliser un regroupement manuel n'est pas réellement envisageable car cela rendrait la tâche extrêmement fastidieuse.

Le deuxième problème est que nous n'avons pas d'exemples (de base de données dont les impacts sont déjà classifiés) qui nous permettraient d'entraîner un modèle et par la suite de

classifier nos impacts grâce à ce modèle. Un apprentissage supervisé, comme avec un réseau de neurones ou une SVM, n'est de ce fait pas envisageable. Nous sommes donc obligés de faire de l'apprentissage non-supervisé.

4.8.6 Etape 6 : Choix de l'algorithme

Une méthode connue pour le clustering est la méthode des *K-Means*. Cet algorithme nécessite de connaître le nombre de clusters. D'après ce que nous avons vu dans la section 2.1, nous avons essayé avec $K = 4$, cependant le résultat obtenu n'était pas cohérent. En effet, aucun cluster n'était constitué seulement d'impacts dont le temps de montée est rapide. De ce fait, nous ne connaissons pas la valeur de K . Nous avons donc choisi d'utiliser 3 algorithmes différents qui sont plus flexibles sur le nombre de clusters et de comparer les résultats obtenus :

- SOM,
- DBSCAN,
- Hierarchical clustering.

Nous avons choisi de tester 3 méthodes différentes car lors de l'apprentissage non-supervisé il est complexe de déterminer si le résultat obtenu est le bon du fait de l'absence d'exemples de comparaison. Utiliser plusieurs méthodes permet donc de comparer les résultats entre eux.

4.8.7 Etape 7 : Application de l'algorithme

Puisque nous avons choisi de tester plusieurs algorithmes, nous allons ici présenter le paramétrage pour chacun d'entre eux.

4.8.7.1 Paramétrage Hierarchical clustering

Le paramètre important est le critère de *linkage* utilisé. Le critère de *linkage* détermine la distance à utiliser entre les ensembles d'observation. L'algorithme fusionnera les paires de clusters qui minimisent ce critère.

Pour déterminer ce critère, nous avons dû tester les différentes possibilités. Nous avons choisi le critère de *linkage* qui nous a permis de minimiser la distance moyenne entre chaque observation et le centroïde du cluster auquel cette observation appartient. De ce fait, nous avons sélectionné la méthode de *linkage "ward"*, qui minimise la variance des *clusters* fusionnés. La formule est la suivante :

$$d(u, v) = \sqrt{\frac{|v| + |s|}{T} d(v, s)^2 + \frac{|v| + |t|}{T} d(v, t)^2 - \frac{|v|}{T} d(s, t)^2}$$

4.8.7.2 Paramétrage SOM

Pour une carte auto adaptative, il y a 3 paramètres principaux à fixer. Le premier est la taille de la carte, autrement dit le nombre de neurones. Vesanto [91] a déterminé une formule heuristique qui permet de calculer ce paramètre : $m = 5 \times \sqrt{n}$ où m correspond au nombre total de neurones et n au nombre d'observations du jeu de données. m permet ensuite de calculer les dimensions x et y de la carte en calculant simplement sa racine carrée. Le second est la propagation de la fonction de voisinage, mais pour ce paramètre il n'y a pas de formule permettant de calculer la valeur optimale. Le dernier paramètre est le taux d'apprentissage initial, pour lequel il n'existe pas non plus de formule de calcul. Pour ces deux derniers paramètres, la valeur doit seulement être comprise dans l'intervalle $]0, x/2]$.

Nous disposons d'environ 3000 lignes dans notre jeu de données. Nous avons donc calculé $m \approx 274$. En calculant la racine carrée de m , nous obtenons environ 16.5. Nous avons donc choisi de prendre une grille de neurones de la taille 16×16 .

Pour ce qui est des deux autres paramètres, nous avons testé 5000 combinaisons possibles. Nous avons fixé $[0.001, 8]$ comme intervalle de test. Pour définir le meilleur paramétrage, nous avons cherché la combinaison qui minimise l'erreur de quantification, qui est calculée comme la distance moyenne entre chaque échantillon d'entrée et son *BMU*. Ceci nous a permis de calculer une propagation de la fonction de voisinage égale à 2.36 et un taux d'apprentissage initial de 5.92.

4.8.7.3 Paramétrage DBSCAN

Le paramètre principal pour DBSCAN est *eps*. Ce paramètre est la distance maximale entre deux échantillons pour que l'un soit considéré comme voisin de l'autre. Le deuxième critère important est *minPts*.

Ici encore, nous avons essayé d'optimiser l'algorithme en testant plusieurs combinaisons de paramètres. L'objectif était également de minimiser la distance moyenne entre chaque échantillon et le centre de leur cluster. De cette manière, nous avons obtenu $eps = 0.012$ et $minPts = 5$.

4.8.8 Etape 8 : Interprétation des résultats

4.8.8.1 Résultats Hierarchical clustering

L'algorithme nous a permis d'obtenir le dendrogramme de la figure 4.37.

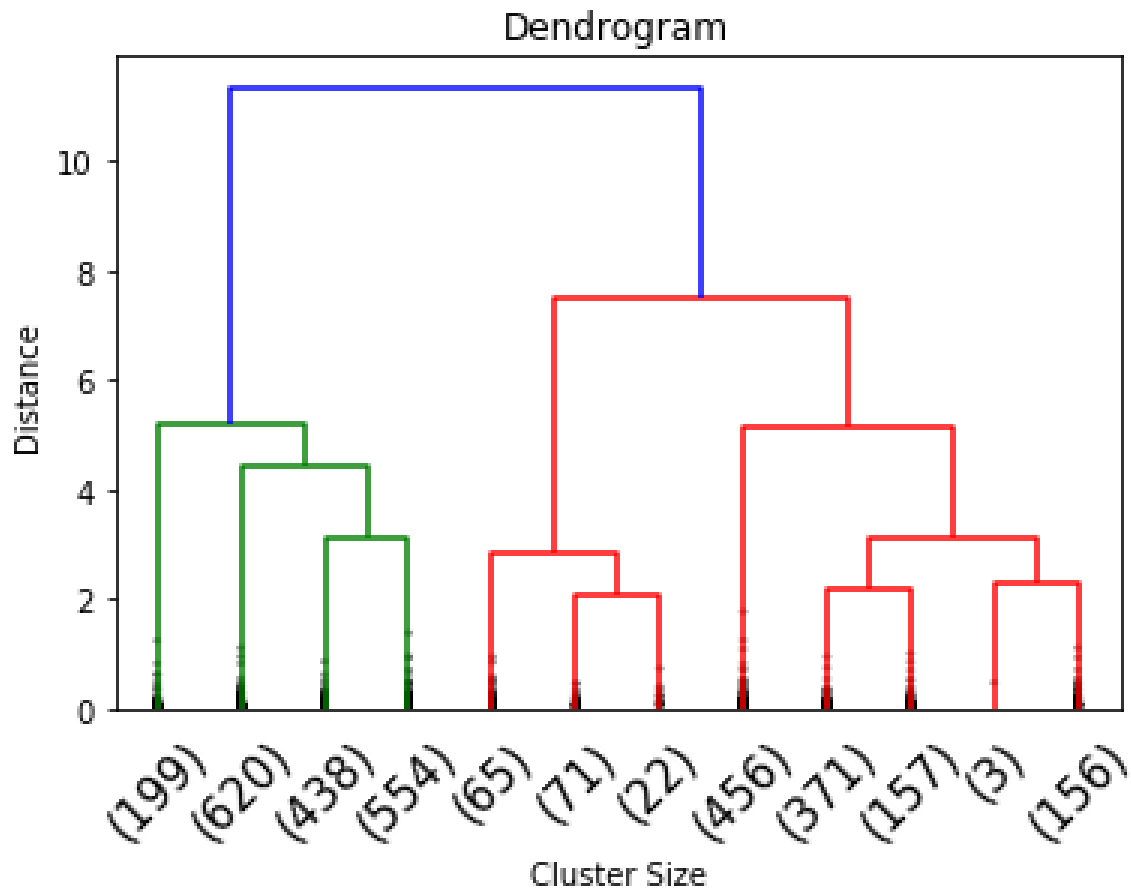


FIGURE 4.37 Dendrogramme représentant le clustering des impacts.

A partir de cette figure, plusieurs interprétations sont possibles, mais elles seraient subjectives. Nous avons opté pour une solution plus objective, qui revient à déterminer le nombre de clusters en minimisant encore une fois la distance moyenne des observations avec leur centroïde. De cette manière, nous avons déterminé que le nombre de clusters optimal est 5.

La distribution des impacts dans les clusters est la suivante :

- Cluster 0 : 1612 impacts,
- Cluster 1 : 687 impacts,
- Cluster 2 : 158 impacts,

- Cluster 3 : 199 impacts,
- Cluster 4 : 456 impacts.

En calculant la moyenne et l'écart-type pour chaque *feature* de chaque cluster, nous avons obtenu le résultat du tableau de la figure 4.38, qui représente les intervalles pour chaque *feature* de chaque cluster.

Crit	Cluster 0		Cluster 1		Cluster 2		Cluster 3		Cluster 4	
RelativeMaxValue	0,281	0,560	0,376	0,930	0,246	0,791	0,135	0,903	0,328	0,575
RiseTimeMax	0,091	0,627	0,296	0,626	0,990	2,452	0,000	0,180	0,738	1,246
DropTimeMax	3,544	5,000	4,050	6,244	3,245	5,118	1,018	2,839	3,566	4,971
Integrale	0,552	1,038	1,046	1,815	0,940	1,937	0,071	0,316	0,847	1,360
Duration	3,973	5,329	4,557	6,734	5,113	7,038	1,079	2,940	4,665	6,013
PressureImpulse	0,039	0,242	0,184	0,427	0,761	1,123	0,000	0,058	0,393	0,631
ImpulseDuration	0,182	1,255	0,592	1,252	2,060	4,785	0,000	0,359	1,476	2,493
	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max

FIGURE 4.38 Tableau représentant les intervalles pour chaque *feature* de chaque cluster du dendrogramme.

La figure 4.39 est un graphique permettant d'appréhender plus facilement les intervalles pour chaque critère de ces clusters.

Nous pouvons observer plusieurs choses sur ce graphique. Premièrement, le critère *RelativeMaxValue* ne semble pas être un critère déterminant pour le clustering de ces données. En effet, les intervalles des différents clusters pour ce critère semblent tous plus ou moins se chevaucher.

Deuxièmement, nous pouvons voir que pour tous les autres critères, le cluster 3 est à chaque fois représenté par l'intervalle minimum. A l'inverse, mis à part pour le *DropTimeMax*, le cluster 2 représente toujours l'intervalle maximum.

En ce qui concerne le *DropTimeMax*, le cluster 3 représente ceux qui sont le plus rapide et le cluster 1 ceux qui sont le plus long. Pour les trois autres clusters, ce critère semble relativement similaire.

Le cluster 0 et le cluster 1 semblent très proches l'un de l'autre. En effet, pour tous les critères nous pouvons voir que les intervalles des critères pour ces clusters se chevauchent ou sont dans la continuité l'un de l'autre, mais qu'à chaque fois le cluster 0 représente la limite inférieure.

De même, nous observons que les intervalles des clusters 2 et 4 se chevauchent pour chacun des critères.

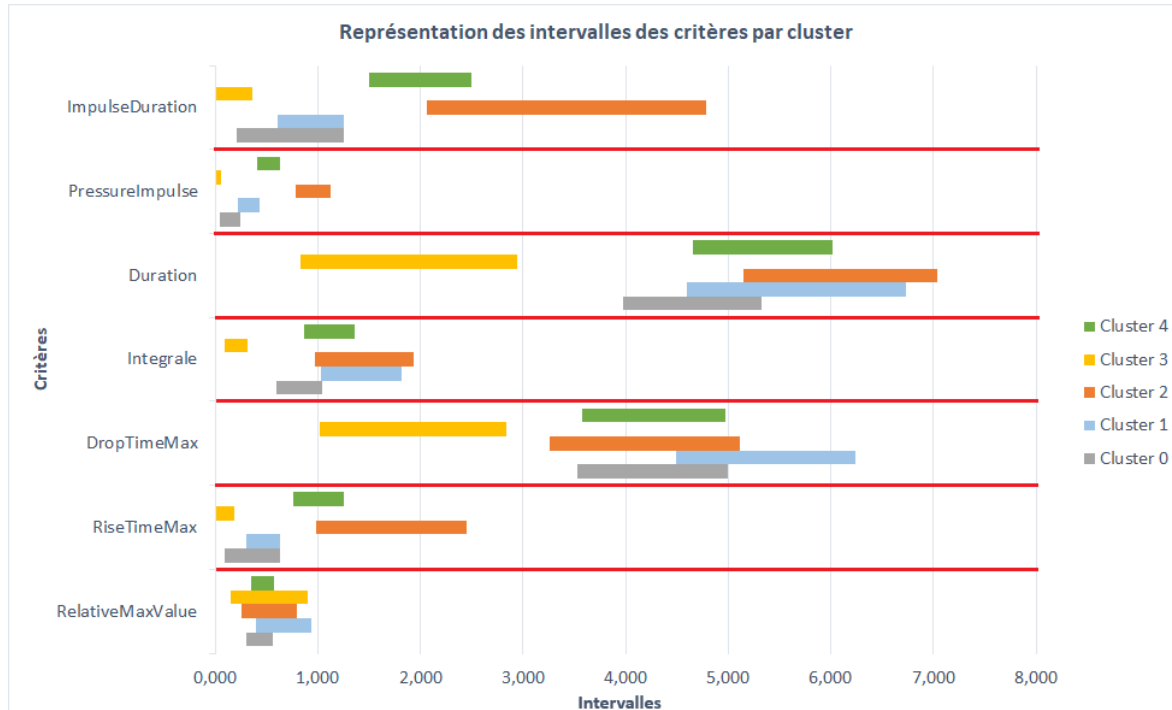


FIGURE 4.39 Graphique représentant les intervalles de chaque critère par cluster obtenu avec Hierarchical clustering.

Malgré le graphique 4.39, il est complexe de bien se représenter les différents clusters. Une autre façon d'examiner ces clusters est de tracer les signaux des impacts. Nous avons donc voulu analyser l'allure des signaux d'impact par cluster, afin de voir si les signaux d'impacts appartenant à un même cluster avaient une allure similaire. Les figures 4.40, 4.41, 4.42, 4.43 et 4.44 montrent des échantillons de signaux d'impact par cluster. Visuellement, nous voyons bien qu'il y a des similitudes intra-cluster. En revanche, il est difficile de différencier les impacts des clusters 0, 1, 2 et 4 de manière visuelle. Pour ce qui est du cluster 3, nous observons que tous les impacts commencent directement par un pic, qui est bien souvent très court.



FIGURE 4.40 Echantillons de signaux d'impact du cluster 0 du dendrogramme.



FIGURE 4.41 Echantillons de signaux d'impact du cluster 1 du dendrogramme.



FIGURE 4.42 Echantillons de signaux d'impact du cluster 2 du dendrogramme.

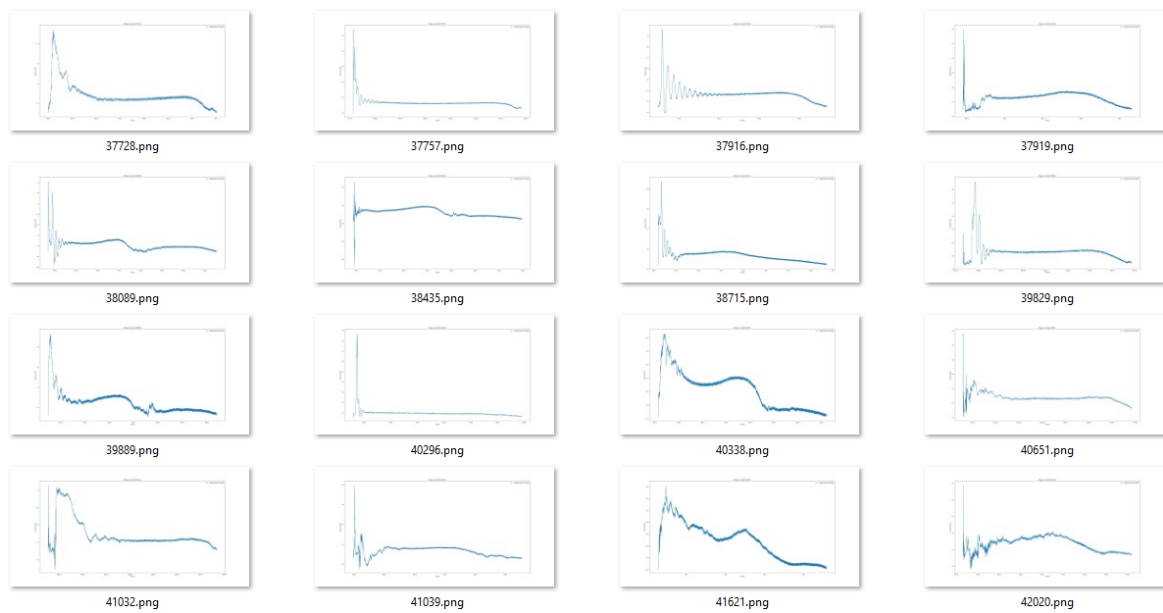


FIGURE 4.43 Echantillons de signaux d'impact du cluster 3 du dendrogramme.



FIGURE 4.44 Echantillons de signaux d'impact du cluster 4 du dendrogramme.

Ces échantillons d'impacts montrent qu'il y a des similitudes entre les signaux d'impact intra-clusters. Cependant, nous pouvons également voir qu'il y a quelques différences au sein d'un même cluster, notamment dans le cluster 3 où l'on peut observer des pics beaucoup plus courts que d'autres. Ces échantillons montrent également des similitudes entre clusters. Pour avoir une vision plus générale des signaux d'impacts par cluster, nous avons fait la moyenne des signaux d'impact par cluster, que nous avons ensuite tracé. Ceci permet d'avoir un signal "type" par cluster. Le résultat est visible sur les figures 4.45, 4.46, 4.47, 4.48, 4.49. En observant ces figures, nous pouvons constater que l'allure des signaux d'impact des clusters 0, 2 et 4 sont très proches. Ce que nous pouvons également observer, c'est que pour le cluster 3, qui contient en partie des impacts avec des pics très rapides (comme nous pouvons le voir dans les échantillons de la figure 4.43), le signal d'impact "type" monte très rapidement mais descend lentement. Ceci montre que dans ce cluster, les impacts avec des pics très rapides sont peu nombreux. Ceci pose un problème car les impacts avec des pics très rapides sont des impacts très intéressants, et le fait qu'ils soient regroupés avec d'autres impacts dont le temps de descente est plus lent masque leur réelle existence.

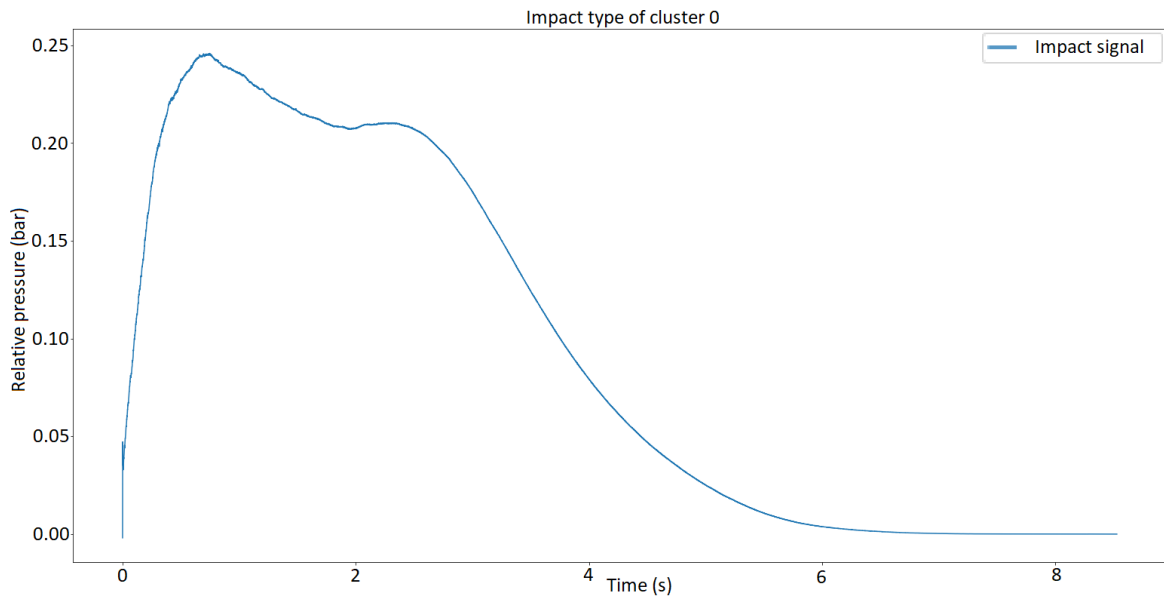


FIGURE 4.45 Signal d'impact type moyenné pour le cluster 0 obtenu avec l'algorithme Hierarchical clustering.

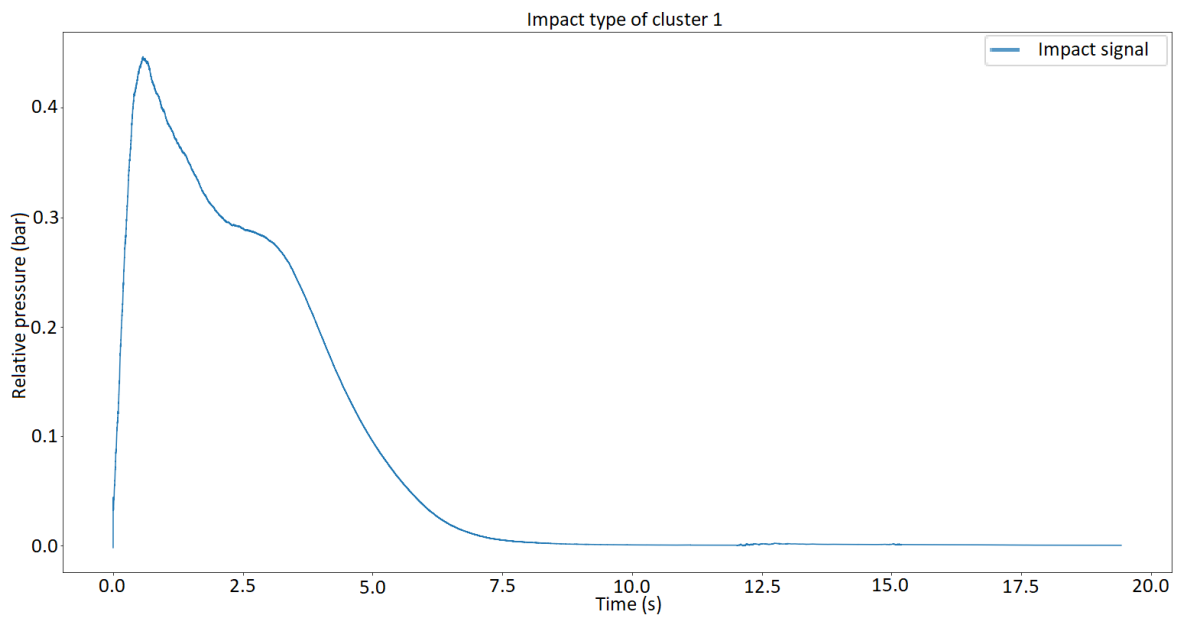


FIGURE 4.46 Signal d'impact type moyenné pour le cluster 1 obtenu avec l'algorithme Hierarchical clustering.

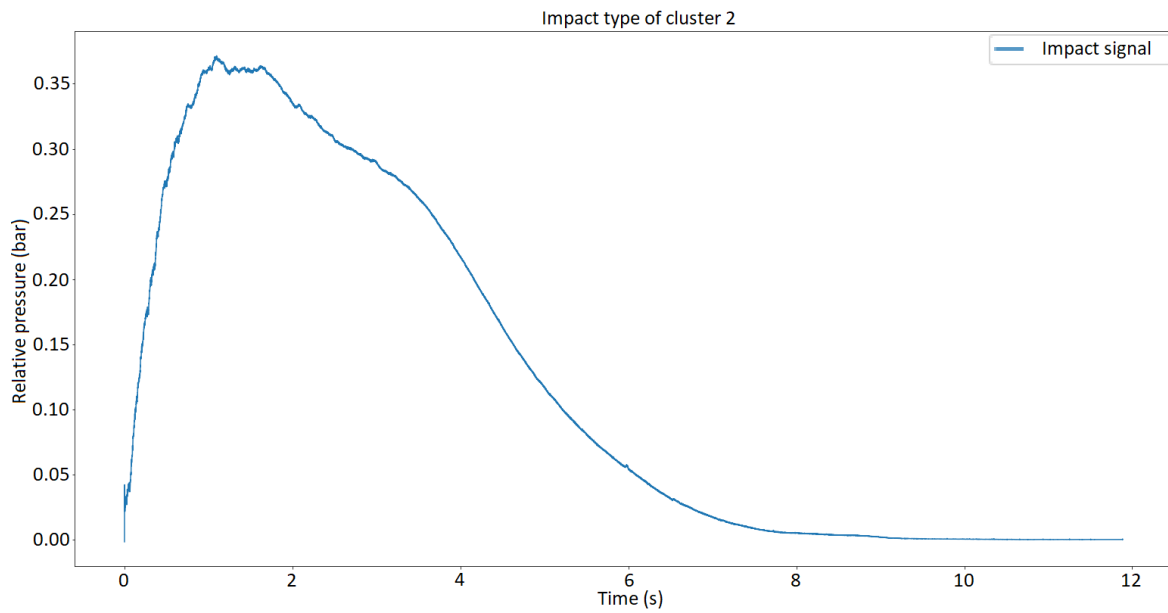


FIGURE 4.47 Signal d'impact type moyenné pour le cluster 2 obtenu avec l'algorithme Hierarchical clustering.

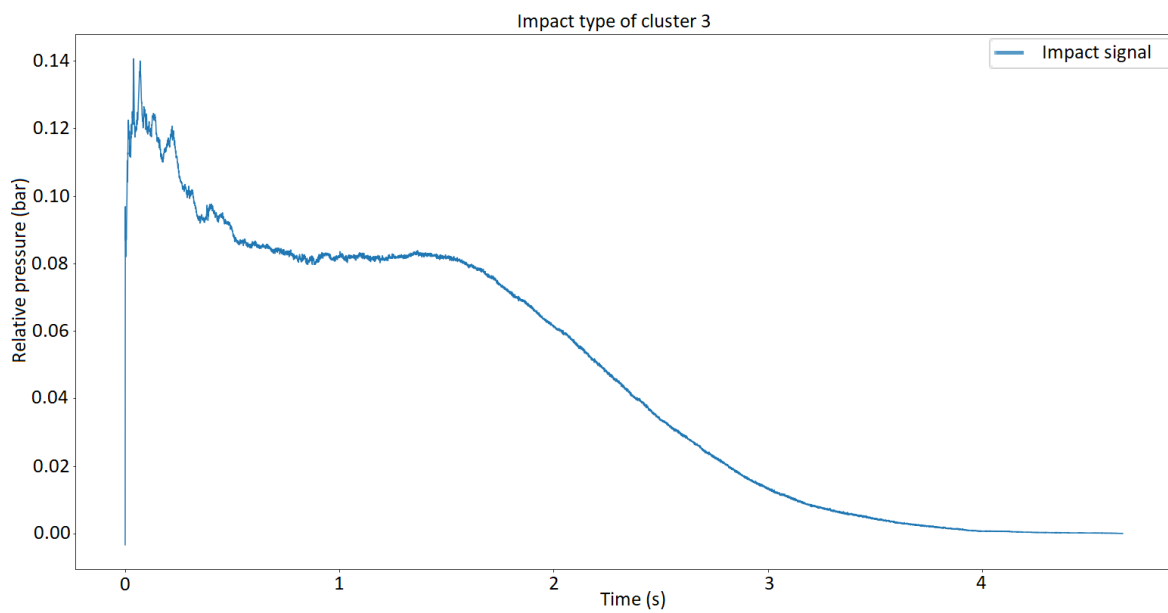


FIGURE 4.48 Signal d'impact type moyenné pour le cluster 3 obtenu avec l'algorithme Hierarchical clustering.

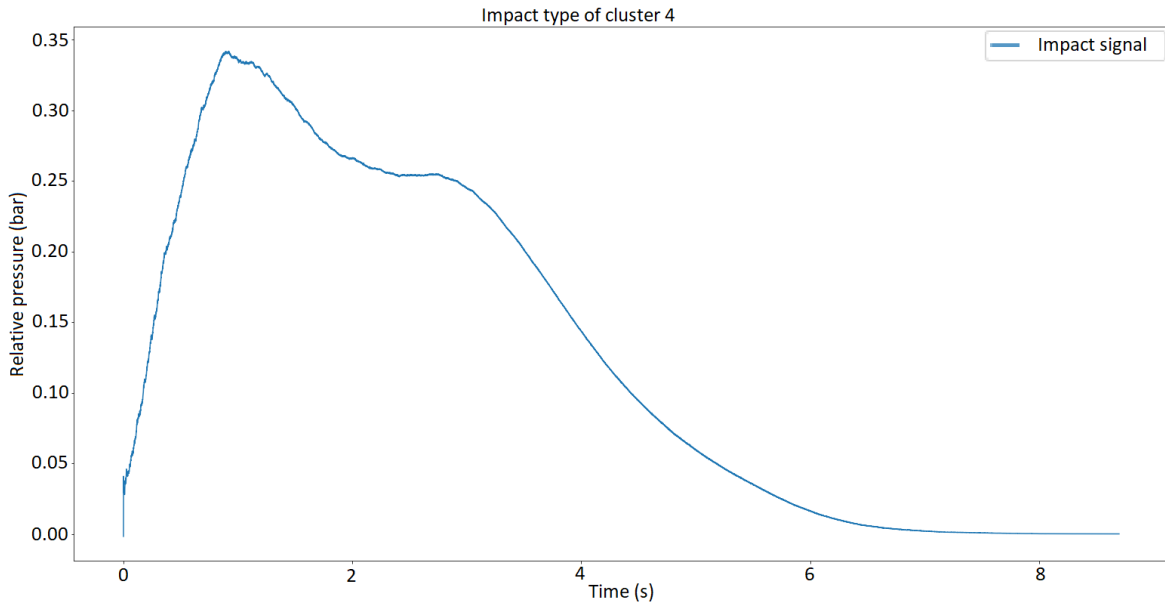


FIGURE 4.49 Signal d'impact type moyenné pour le cluster 4 obtenu avec l'algorithme Hierarchical clustering.

4.8.8.2 Résultats SOM

L'algorithme paramétré tel que décrit dans la section 4.8.7.2 nous a permis d'obtenir la carte de la figure 4.50. Sur cette figure, la couleur d'un neurone correspond à sa distance avec ses voisins. Plus un neurone sera proche de ses voisins, plus il sera blanc. A l'inverse, plus il sera éloigné de ses voisins, plus il sera noir.

Il est très compliqué d'interpréter une telle carte afin d'établir des clusters. Nous pouvons voir en bas à gauche de la carte, que tous les neurones sont très proches les uns des autres, et de ce fait appartiennent probablement au même cluster. Les quatre neurones très noirs sont quant à eux très loin de leurs voisins. Nous pouvons donc imaginer qu'ils représentent chacun un cluster différent. Le problème de cette carte est d'interpréter tous les autres neurones. Est-ce qu'ils appartiennent à un des clusters mentionnés juste avant? Ou forment-ils de nouveaux clusters? Ces questions sont complexes et il est difficile d'y répondre.

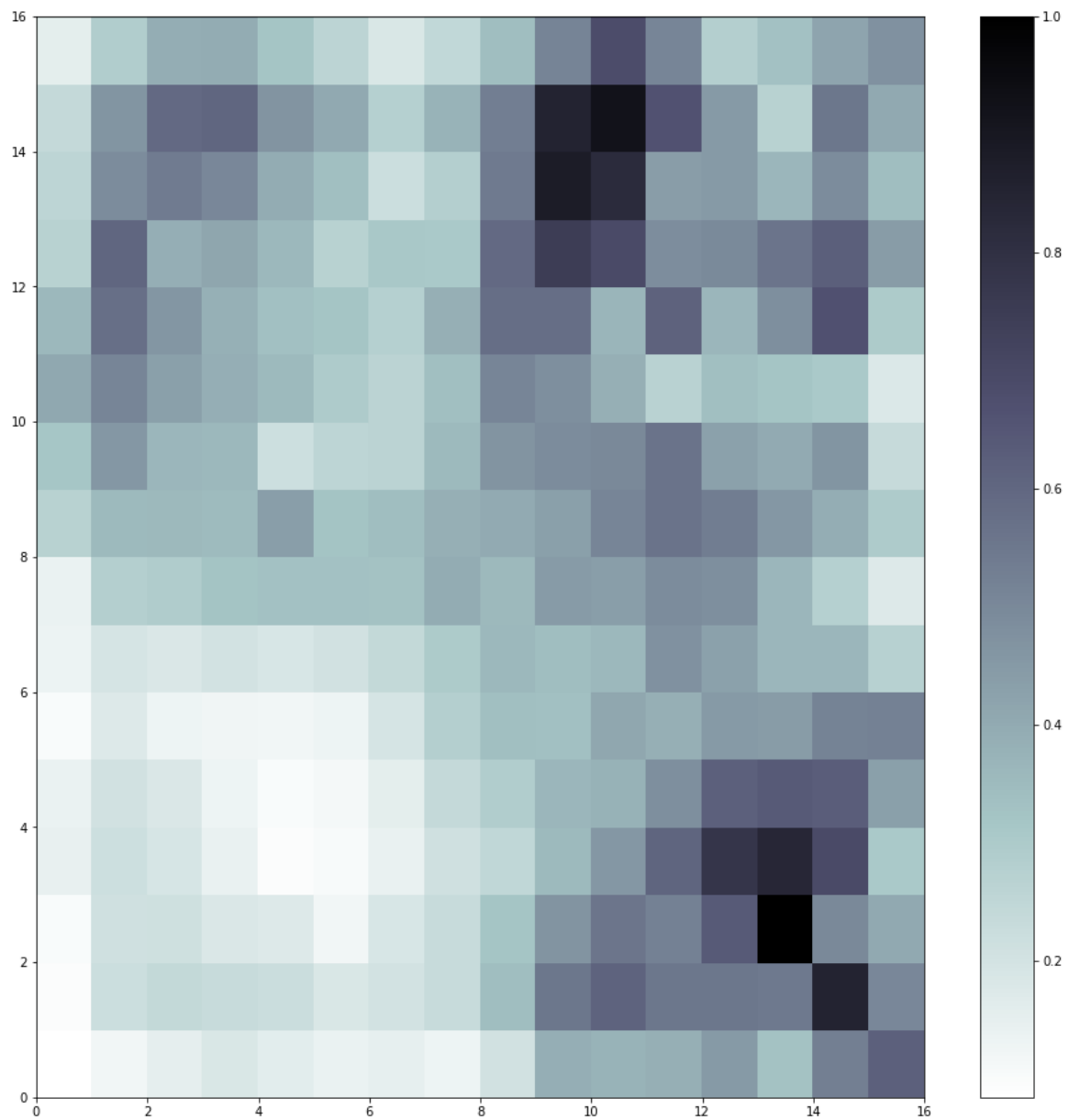


FIGURE 4.50 Carte de dimension 16 par 16 obtenue avec l'algorithme SOM.

De ce fait, nous avons voulu faire la même chose, mais en réduisant considérablement la taille de la carte. Nous avons donc choisi une carte de taille 4 par 4. Les paramètres optimaux que nous avons obtenus pour cette taille de carte sont les suivantes :

- Propagation de la fonction de voisinage = 0.75
- Taux d'apprentissage initial de 0.88

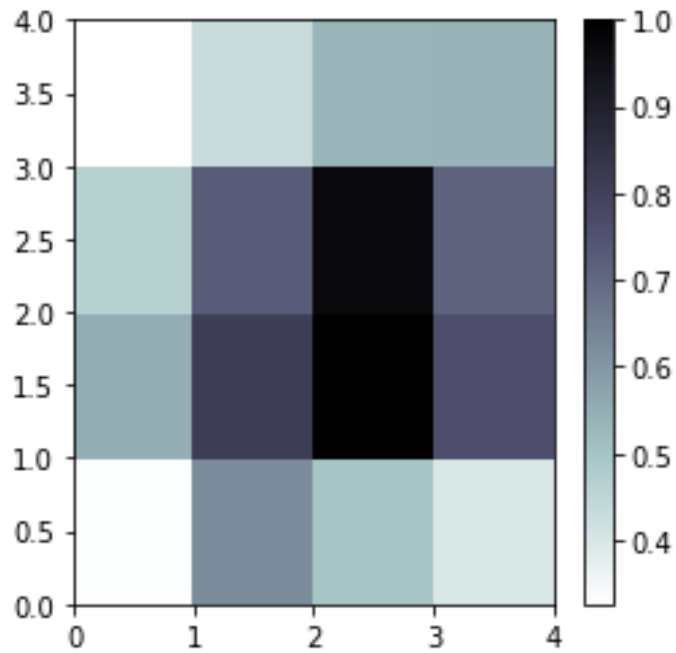


FIGURE 4.51 Carte de dimension 4 par 4 obtenue avec l'algorithme SOM.

Sur cette carte contenant moins de neurones, nous pouvons plus facilement essayer d'interpréter le résultat et donc déterminer le nombre de clusters. Nous avons tenté de faire cela, en fusionnant les neurones de la manière montrée sur la figure 4.52. Ce regroupement de neurones est subjectif, et d'autres interprétations sont possibles.

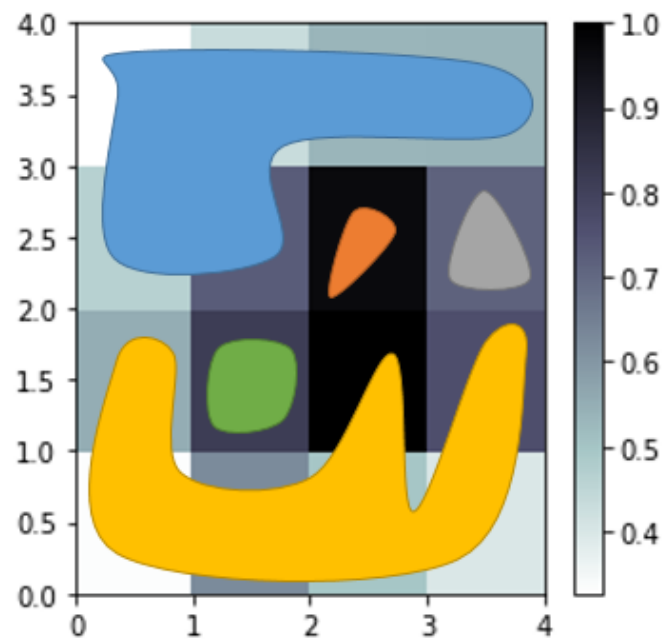


FIGURE 4.52 Clusters déterminés par l'analyse de la carte de dimension 4 par 4 obtenue avec l'algorithme SOM.

La distribution des impacts dans les clusters est la suivante :

- Cluster 0 : 1286 impacts,
- Cluster 1 : 197 impacts,
- Cluster 2 : 1154 impacts,
- Cluster 3 : 306 impacts,
- Cluster 4 : 169 impacts.

En calculant la moyenne et l'écart-type pour chaque *feature* de chaque cluster, nous avons obtenus le résultat du tableau de la figure 4.53, qui représente les intervalles pour chaque *feature* de chaque cluster.

Crit	Cluster 0		Cluster 1		Cluster 2		Cluster 3		Cluster 4	
RelativeMaxValue	0,276	0,781	0,481	0,685	0,244	0,709	0,310	0,410	0,298	0,453
RiseTimeMax	0,244	1,334	0,392	0,637	0,032	0,314	0,508	0,737	0,572	0,955
DropTimeMax	3,819	5,964	3,939	4,714	2,770	4,984	3,793	4,358	2,909	3,596
Integrale	0,889	1,709	0,986	1,264	0,350	1,036	0,661	0,877	0,505	0,760
Duration	4,831	6,661	4,520	5,233	2,918	5,207	4,424	5,041	3,665	4,431
PressureImpulse	0,170	0,696	0,226	0,382	0,008	0,143	0,175	0,305	0,179	0,370
ImpulseDuration	0,506	2,646	0,784	1,274	0,063	0,627	1,017	1,475	1,144	1,911
	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max

FIGURE 4.53 Tableau représentant les intervalles pour chaque *feature* de chaque cluster du SOM.

La figure 4.54 est un graphique permettant d'appréhender plus facilement les intervalles pour chaque critère de ces clusters. Comme pour le clustering réalisé avec l'algorithme *Hierarchical clustering*, le critère *RelativeMaxValue* ne semble pas être déterminant. Ce que nous pouvons observer ici, c'est que pour le cluster 0 et le cluster 2, les intervalles sont souvent grands, ce qui veut dire que pour un critère donné, les différentes valeurs de ce critère pour les impacts de ces clusters sont hétérogènes. Ceci nous amène à nous interroger sur la validité de ce clustering.

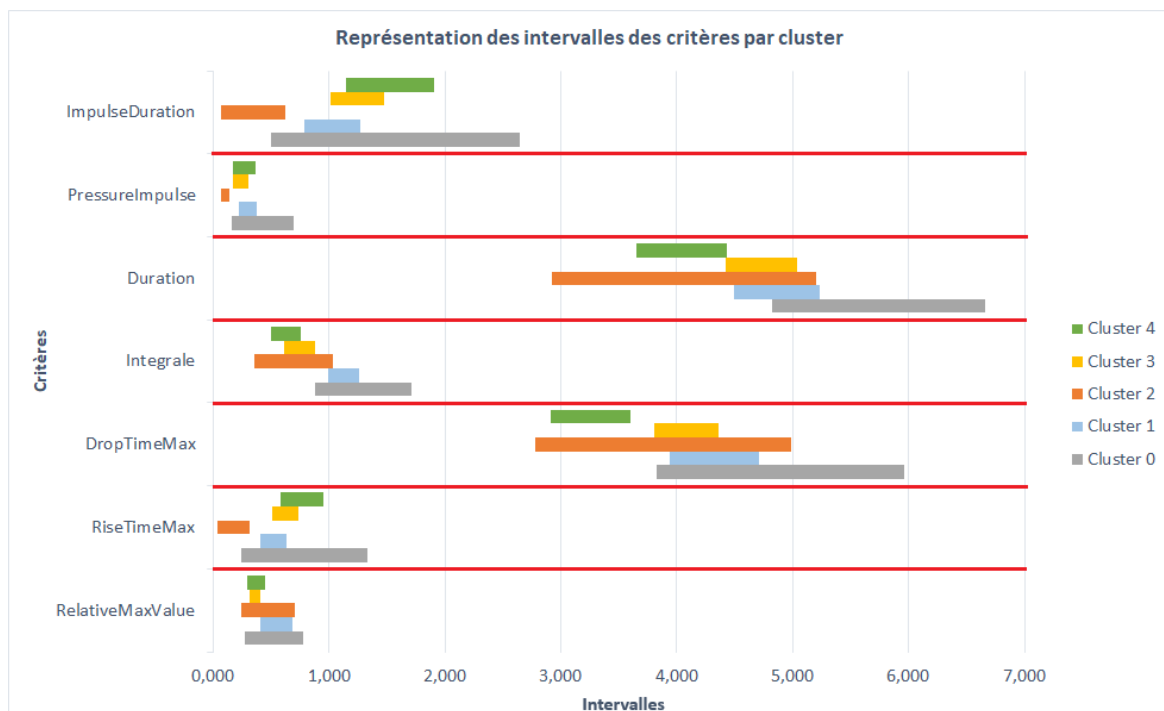


FIGURE 4.54 Graphique représentant les intervalles de chaque critère par cluster obtenu avec SOM.

Pour confirmer nos doutes nous avons tracé les signaux des impacts par cluster. Les figures 4.55, 4.56, 4.57, 4.58 et 4.59 montrent des échantillons de signaux d'impact par cluster. En effet, en observant ces échantillons nous voyons que les différences entre clusters équées que celles des clusters obtenus avec l'algorithme *Hierarchical clustering*.

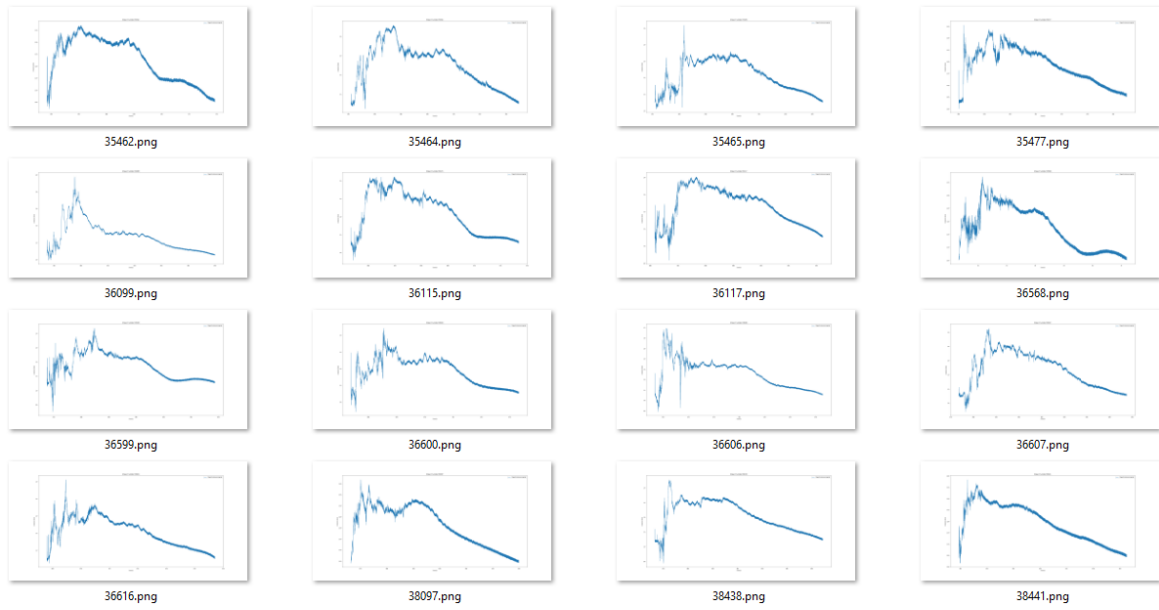


FIGURE 4.55 Echantillons de signaux d'impact du cluster 0 du SOM.

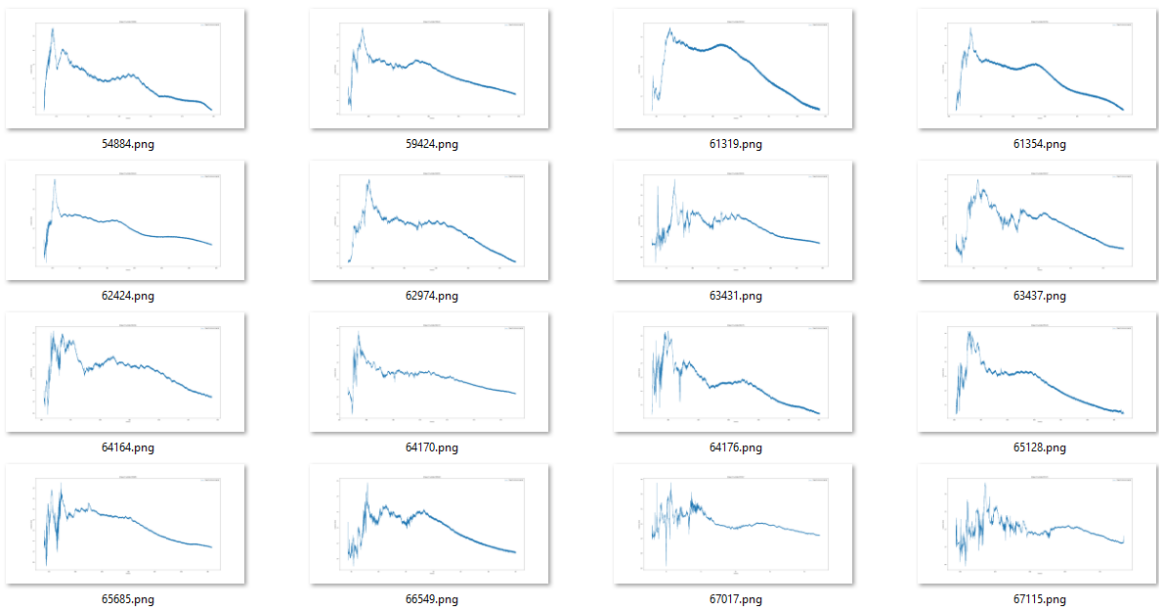


FIGURE 4.56 Echantillons de signaux d'impact du cluster 1 du SOM.



FIGURE 4.57 Echantillons de signaux d'impact du cluster 2 du SOM.

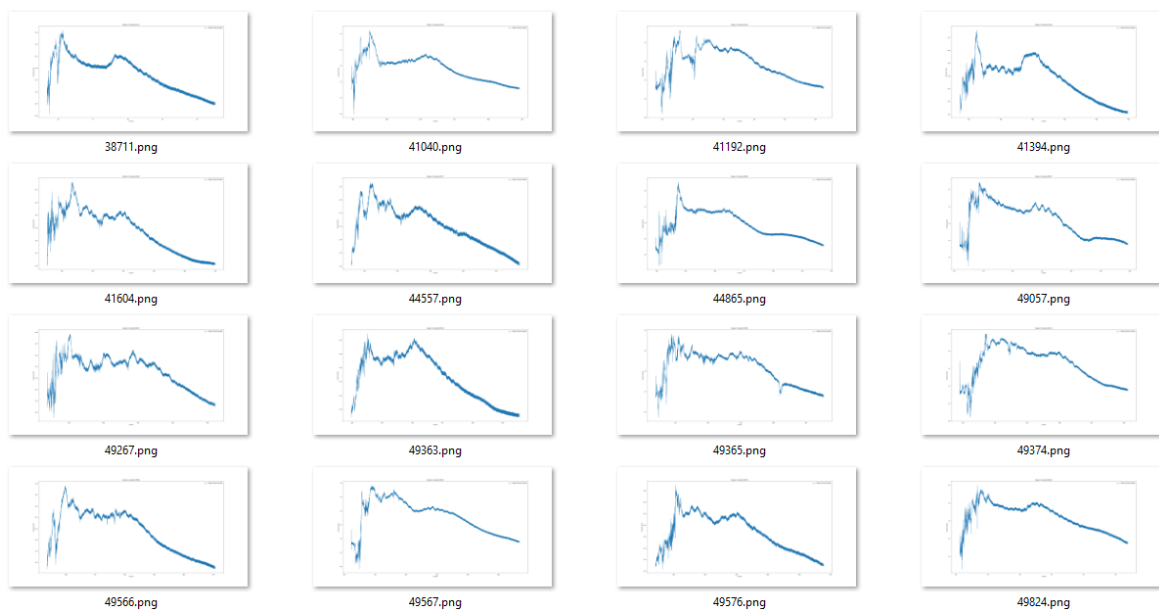


FIGURE 4.58 Echantillons de signaux d'impact du cluster 3 du SOM.

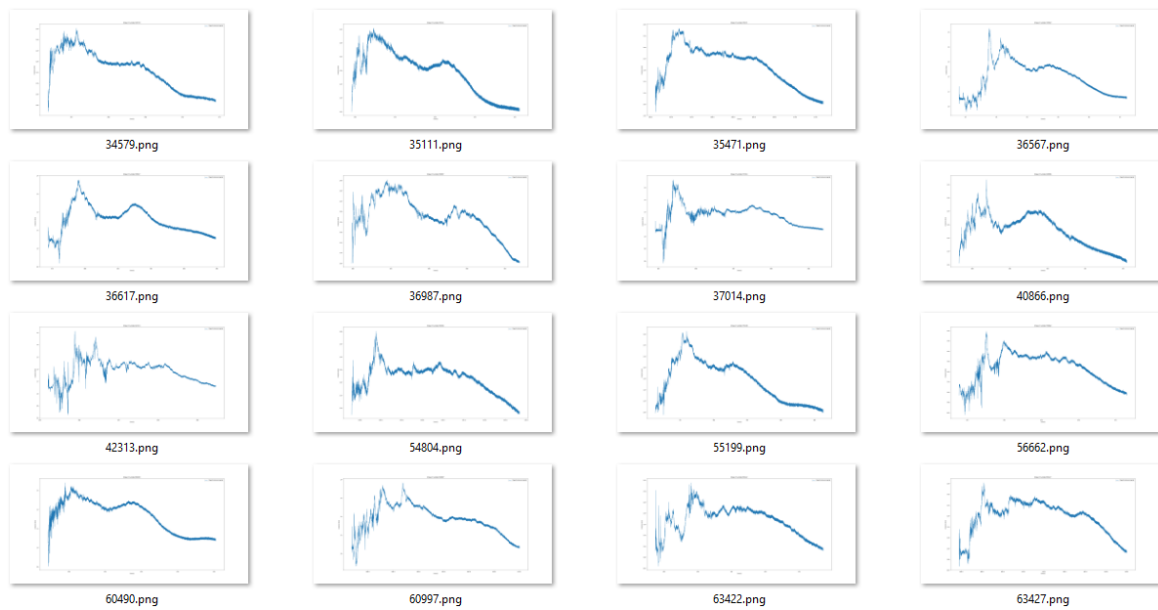


FIGURE 4.59 Echantillons de signaux d'impact du cluster 4 du SOM.

Afin de suivre le même procédé que dans l'approche précédente, nous avons fait une moyenne des impacts par cluster afin d'avoir un impact "type" pour chaque cluster. Ces impacts sont visibles sur les figures 4.60, 4.61, 4.62, 4.63, 4.64. Ce qui est observable sur ces figures est qu'en effet les différences entre types d'impacts sont moins marquées et qu'en plus il n'y a pas de cluster ayant un signal d'impact avec un temps de montée très rapide.

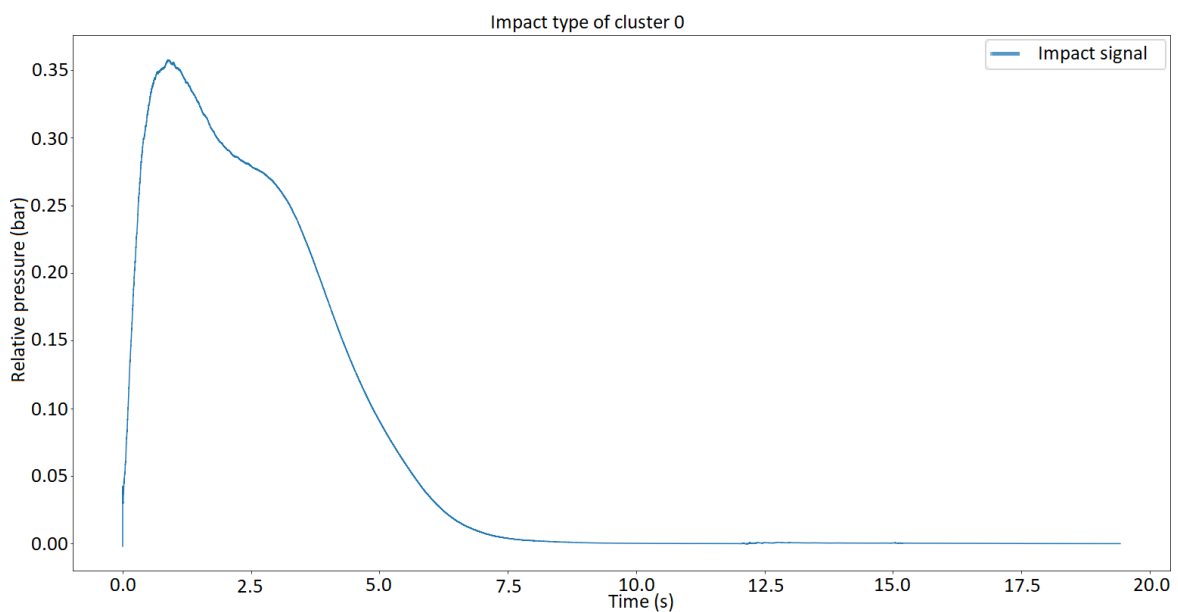


FIGURE 4.60 Signal d'impact type moyenné pour le cluster 0 obtenu avec l'algorithme SOM.

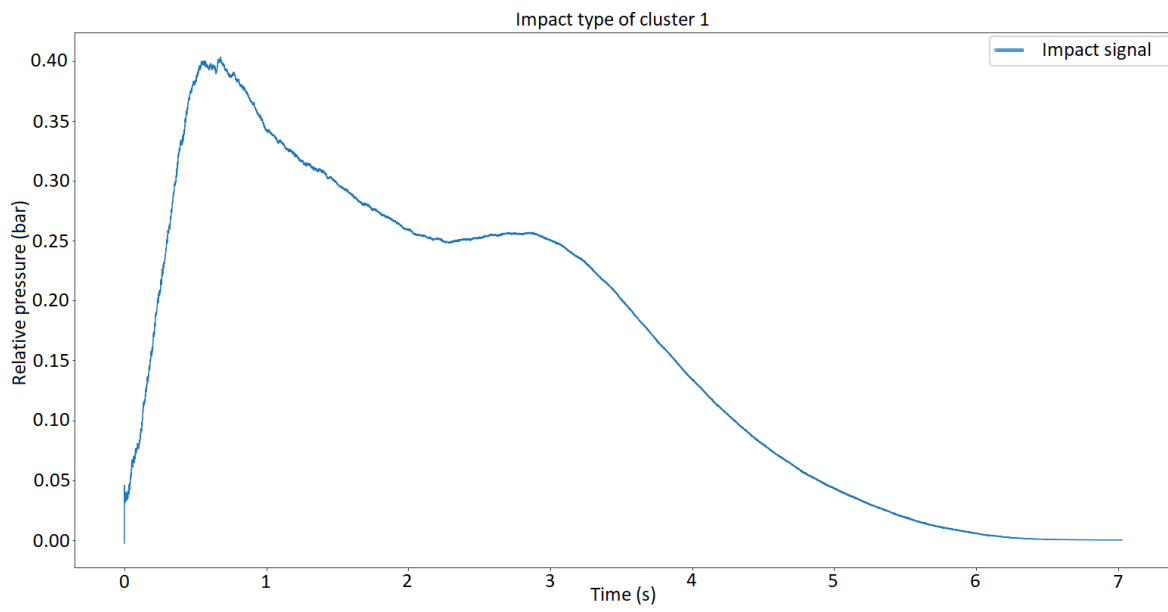


FIGURE 4.61 Signal d'impact type moyenné pour le cluster 1 obtenu avec l'algorithme SOM.

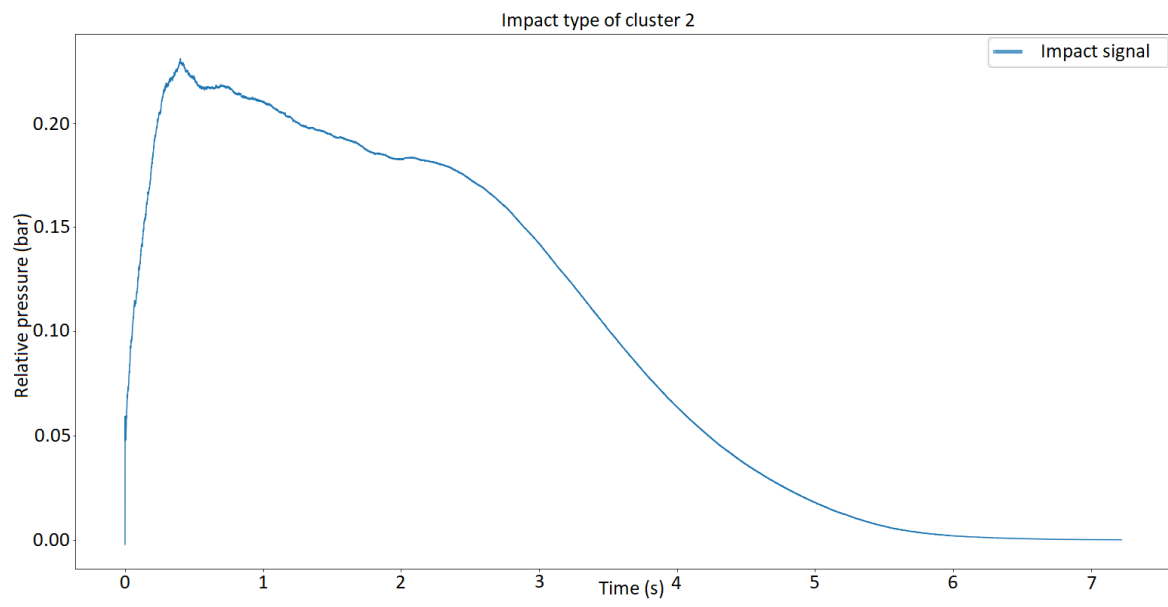


FIGURE 4.62 Signal d'impact type moyenné pour le cluster 2 obtenu avec l'algorithme SOM.

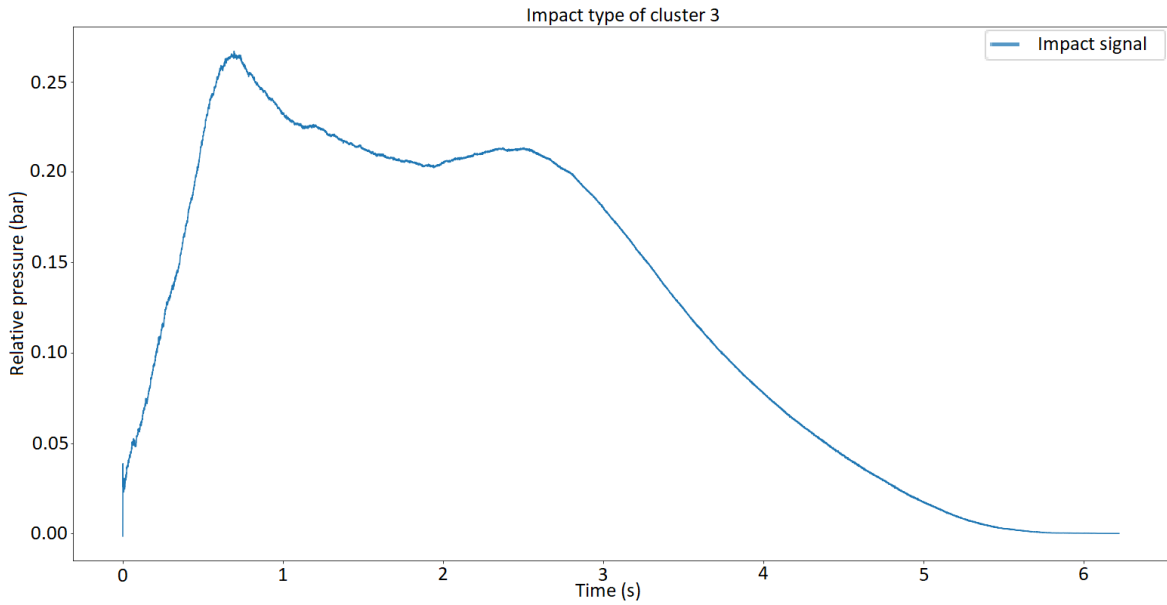


FIGURE 4.63 Signal d'impact type moyenné pour le cluster 3 obtenu avec l'algorithme SOM.

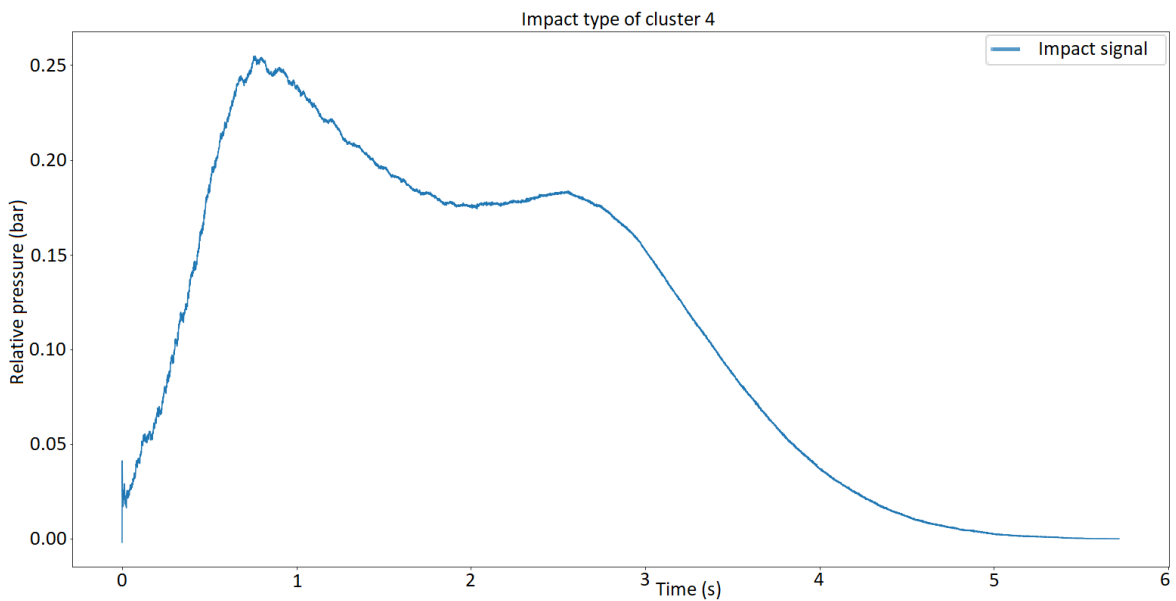


FIGURE 4.64 Signal d'impact type moyenné pour le cluster 4 obtenu avec l'algorithme SOM.

4.8.8.3 Résultats DBSCAN

Cet algorithme n'a pas donné de résultats exploitables. En effet, le paramétrage présenté dans la partie 4.8.7.3 nous a permis d'obtenir la figure 4.65. Sur cette figure nous pouvons observer un seul cluster d'impact (points en rouge) et plusieurs impacts considérés comme hors cluster (points en noir). Après ce résultat, nous avons essayé plusieurs paramétrages

différents afin d'obtenir un meilleur clustering, cependant aucun résultat obtenu n'était pertinent.

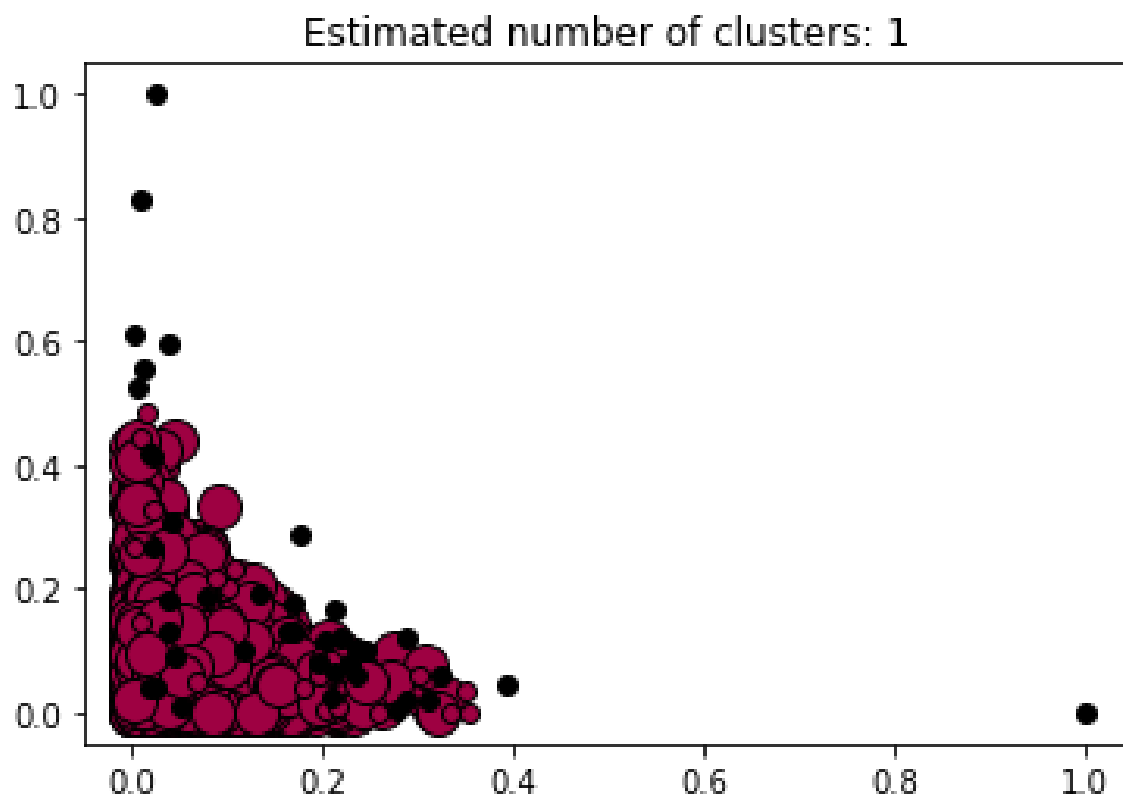


FIGURE 4.65 Clustering obtenu avec l'algorithme DBSCAN.

4.8.9 Comparaison des clusters

Il y a un intérêt à comparer les clusters obtenus avec les algorithmes SOM et *Hierarchical clustering* (HC pour la suite de cette sous-section). En effet, cela permettrait de savoir si les clusters établis par les deux algorithmes sont relativement semblables ou au contraire complètement différents. Pour cela, nous avons choisi de comparer les impacts inclus dans chaque cluster obtenu d'un des algorithmes avec les impacts contenus dans les clusters obtenus avec l'autre algorithme. Cela a donné le tableau présenté sur la figure 4.66.

HC Clusters id	Nombre d'impacts total	SOM Clusters id	Nombre d'impacts total	Nombre d'impacts en commun	Pourcentage en commun par rapport au cluster HC	Pourcentage en commun par rapport au cluster SOM
0	1612	0	1286	236	14,64	18,35
		1	197	21	1,30	10,66
		2	1154	915	56,76	79,29
		3	306	280	17,37	91,50
		4	169	160	9,93	94,67
1	687	0	1286	497	72,34	38,65
		1	197	142	20,67	72,08
		2	1154	41	5,97	3,55
		3	306	7	1,02	2,29
		4	169	0	0,00	0,00
2	158	0	1286	158	100,00	12,29
		1	197	0	0,00	0,00
		2	1154	0	0,00	0,00
		3	306	0	0,00	0,00
		4	169	0	0,00	0,00
3	199	0	1286	0	0,00	0,00
		1	197	0	0,00	0,00
		2	1154	198	99,50	17,16
		3	306	0	0,00	0,00
		4	169	1	0,50	0,59
4	456	0	1286	395	86,62	30,72
		1	197	34	7,46	17,26
		2	1154	0	0,00	0,00
		3	306	19	4,17	6,21
		4	169	8	1,75	4,73

FIGURE 4.66 Comparaison des impacts en commun entre les clusters des deux algorithmes.

Ce tableau nous permet de visualiser les résultats suivants :

Cluster HC0 : Le cluster HC0 compte 1612 impacts. Le nombre d'impacts en commun entre les clusters HC0 et SOM2 représente plus de la moitié (56%) des impacts du cluster HC0. Les 44% restants du cluster HC0 sont répartis entre les autres clusters de SOM.

Cluster HC1 : Le cluster HC1 compte 687 impacts. Le nombre d'impacts en commun entre les clusters HC1 et SOM0 représente 72% des impacts de HC1. Le nombre d'impacts en commun entre les clusters HC1 et SOM1 représente 20% des impacts de HC1. Le cluster HC1 est donc un mixte entre le cluster SOM0 et le cluster SOM1, avec une tendance plus marquée par le cluster SOM0.

Cluster HC2 : Le cluster HC2 compte 158 impacts. Le nombre d'impacts en commun entre les clusters HC2 et SOM0 représente 100% des impacts de HC2. Le cluster HC2 est donc complètement inclus dans le cluster SOM0.

Cluster HC3 : Le cluster HC3 compte 199 impacts. Le nombre d'impacts en commun entre les clusters HC3 et SOM2 représente 99.5% des impacts de HC3. Le cluster HC3 est donc complètement inclus dans le cluster SOM2, à l'exception d'un impact sur 199.

Cluster HC4 : Le cluster HC4 compte 456 impacts. Le nombre d'impacts en commun entre les clusters HC4 et SOM0 représente plus de 86% des impacts du cluster HC4.

Cluster SOM0 : Le cluster SOM0 compte 1286 impacts. Ce cluster a des impacts en commun avec 4 des 5 clusters de HC. En effet, le nombre d'impacts en commun entre ce cluster et ceux obtenus avec HC est réparti comme suit : - HC 0 : 18% - HC 1 : 38% - HC 2 : 12% - HC 3 : 0% - HC 4 : 30%

Cluster SOM1 : Le cluster SOM1 compte 197 impacts. Le nombre d'impacts en commun entre les clusters SOM1 et HC1 représente 72% des impacts du cluster SOM1. Les 28% restants du cluster SOM1 sont répartis entre les clusters HC0 et HC4.

Cluster SOM2 : Le cluster SOM2 compte 1154 impacts. Le nombre d'impacts en commun entre les clusters SOM2 et HC0 représente presque 80% des impacts du cluster SOM2. Les 20% restants du cluster SOM2 sont principalement inclus dans le cluster HC3.

Cluster SOM3 : Le cluster SOM3 compte 306 impacts. Le nombre d'impacts en commun entre les clusters SOM3 et HC0 représente presque 92% des impacts du cluster SOM3. Le cluster SOM3 est donc quasiment totalement inclus dans le cluster HC0.

Cluster SOM4 : Le cluster SOM4 compte 169 impacts. Le nombre d'impacts en commun entre les clusters SOM4 et HC0 représente presque 94% des impacts du cluster SOM4. Le cluster SOM4 est donc lui aussi quasiment totalement inclus dans le cluster HC0.

Ce que nous pouvons conclure de ce tableau est que les clusters obtenus avec chaque algorithme de clustering sont différents, bien que souvent les impacts d'un cluster donné pour un algorithme donné sont en majorité inclus dans un cluster précis de l'autre algorithme.

Nous avons obtenu deux résultats de clustering qui diffèrent et qui de plus sont difficilement comparables. Nous pouvons néanmoins noter quelques différences majeures entre ces deux résultats de clustering :

1. Le résultat qui découle de l'algorithme SOM est beaucoup plus sujet à l'interprétation subjective que le résultat obtenu grâce à l'algorithme *Hierarchical clustering*. En effet, pour ce dernier nous avons déterminé le nombre de clusters de manière objective en minimisant la distance moyenne entre les observations et le centroïde de leur cluster.
2. Les différences entre les clusters créés par l'algorithme SOM sont moins marquées que les différences entre les clusters créés par l'algorithme *Hierarchical clustering*.

3. Dans les clusters de l'algorithme SOM, il n'y a aucun cluster qui est seulement constitué d'impacts dont le temps de montée est rapide, alors que ce type d'impacts est souvent observable.

Cependant, un point commun important est le nombre de clusters obtenus qui est similaire dans les deux algorithmes SOM et *Hierarchical clustering*. Ce nombre de clusters est de 5.

4.8.10 K-Means

Les résultats obtenus présentés précédemment ont permis de déterminer un nombre de clusters égal à 5. Notre première intuition, comme expliqué dans la sous-section 4.8.6, a été d'utiliser ces algorithmes là car ils ne nécessitent pas un nombre de clusters prédéterminé. Mais maintenant que nous avons a priori une valeur $K = 5$, alors il est tout à fait justifiable d'utiliser l'algorithme K-Means, qui est un algorithme de clustering réputé. Cela permettra de comparer les résultats obtenus avec cet algorithme aux résultats obtenus avec les autres algorithmes.

Nous avons donc utilisé K-Means pour regrouper les impacts en 5 clusters. La distribution des impacts dans les clusters est la suivante :

- Cluster 0 : 235 impacts,
- Cluster 1 : 415 impacts,
- Cluster 2 : 1263 impacts,
- Cluster 3 : 449 impacts,
- Cluster 4 : 750 impacts.

En calculant la moyenne et l'écart-type pour chaque *feature* de chaque cluster, nous avons obtenu le résultat du tableau de la figure 4.67, qui représente les intervalles pour chaque *feature* de chaque cluster. En comparant ce tableau avec le tableau identique obtenu pour les clusters de l'algorithme *Hierarchical Clustering* nous pouvons noter que chaque cluster obtenu avec K-Means est très proche d'un cluster obtenu avec *Hierarchical Clustering*. En effet, a priori nous pouvons voir des similitudes entre les clusters suivants (HC pour *Hierarchical clustering* et KM pour K-Means) :

- HC0 et KM2,
- HC1 et KM3,
- HC2 et KM0,
- HC3 et KM1,

— HC4 et KM4.

Crit	Cluster 0		Cluster 1		Cluster 2		Cluster 3		Cluster 4	
	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max
RelativeMaxValue	0,277	0,700	0,173	0,748	0,293	0,587	0,473	1,049	0,306	0,533
RiseTimeMax	0,949	2,213	0,000	0,251	0,136	0,499	0,239	0,605	0,615	1,061
DropTimeMax	3,227	4,970	1,707	3,889	3,857	5,220	4,217	6,624	3,511	4,819
Integrale	0,926	1,723	0,152	0,512	0,672	1,101	1,324	2,018	0,735	1,172
Duration	4,873	6,787	1,805	4,021	4,229	5,521	4,666	7,097	4,405	5,711
PressureImpulse	0,670	1,052	0,000	0,080	0,058	0,208	0,176	0,489	0,280	0,491
ImpulseDuration	1,965	4,333	0,000	0,502	0,271	0,999	0,477	1,210	1,229	2,122

FIGURE 4.67 Tableau représentant les intervalles pour chaque *feature* de chaque cluster de K-Means.

La figure ?? est un graphique permettant d'appréhender plus facilement les intervalles pour chaque critère de ces clusters.

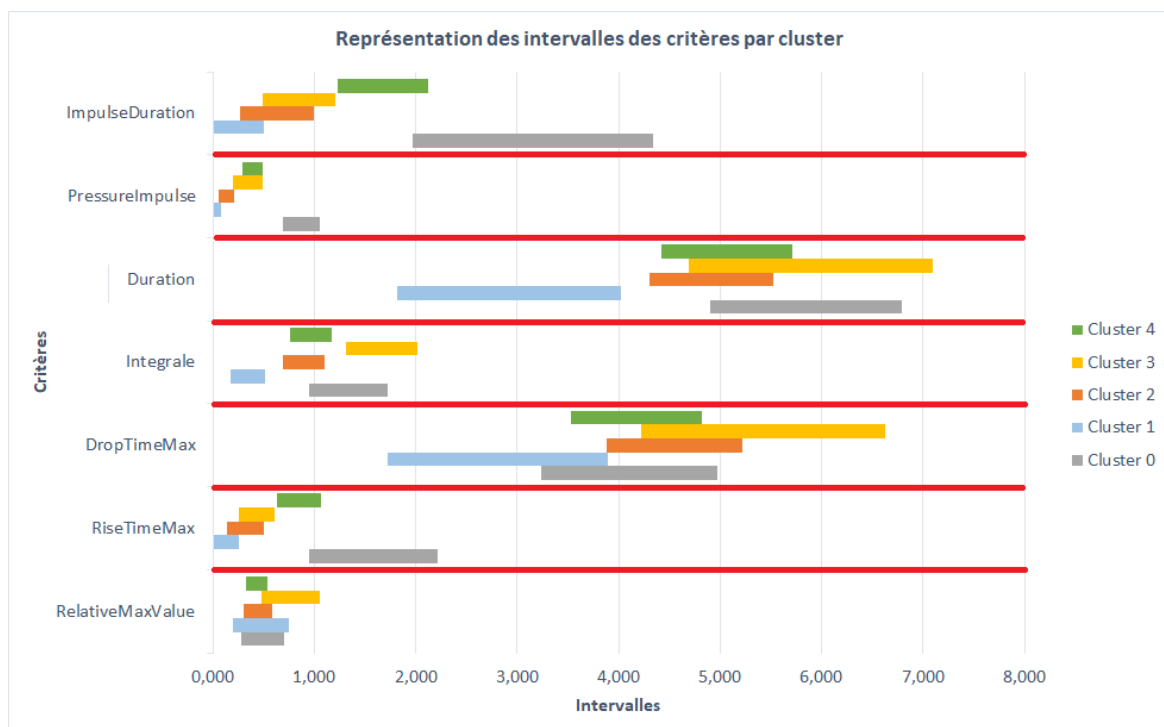


FIGURE 4.68 Graphique représentant les intervalles de chaque critère par cluster obtenu avec K-Means.

Les figures 4.69, 4.70, 4.71, 4.72 et 4.73 montrent chacune un échantillon des signaux d'impacts de chaque cluster.



FIGURE 4.69 Echantillons de signaux d'impact du cluster 0 de K-Means.

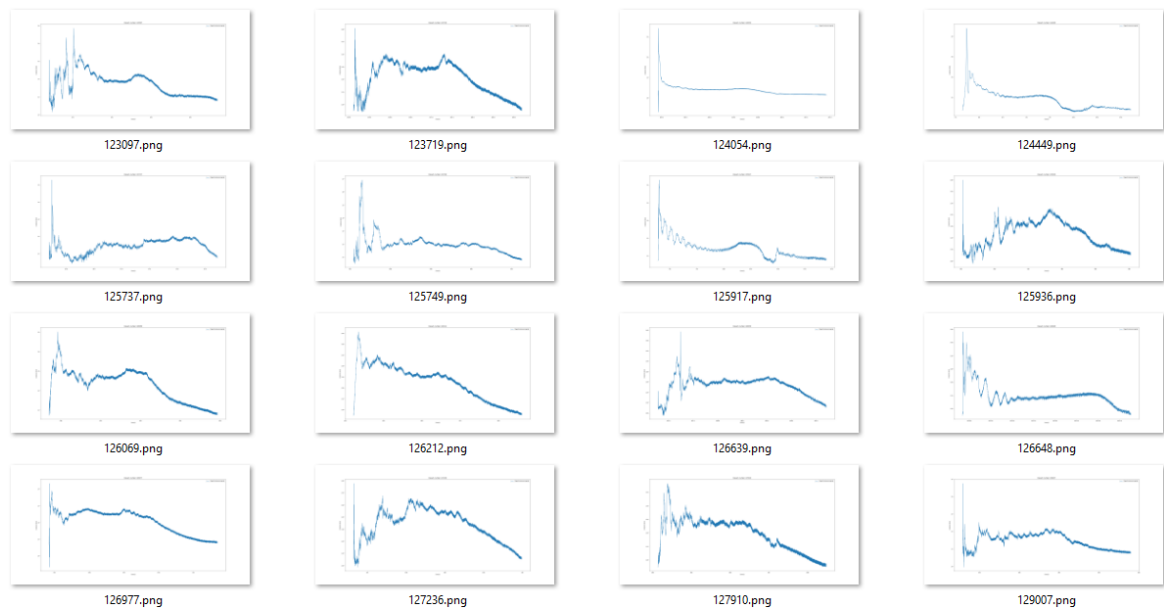


FIGURE 4.70 Echantillons de signaux d'impact du cluster 1 de K-Means.



FIGURE 4.71 Echantillons de signaux d'impact du cluster 2 de K-Means.



FIGURE 4.72 Echantillons de signaux d'impact du cluster 3 de K-Means.

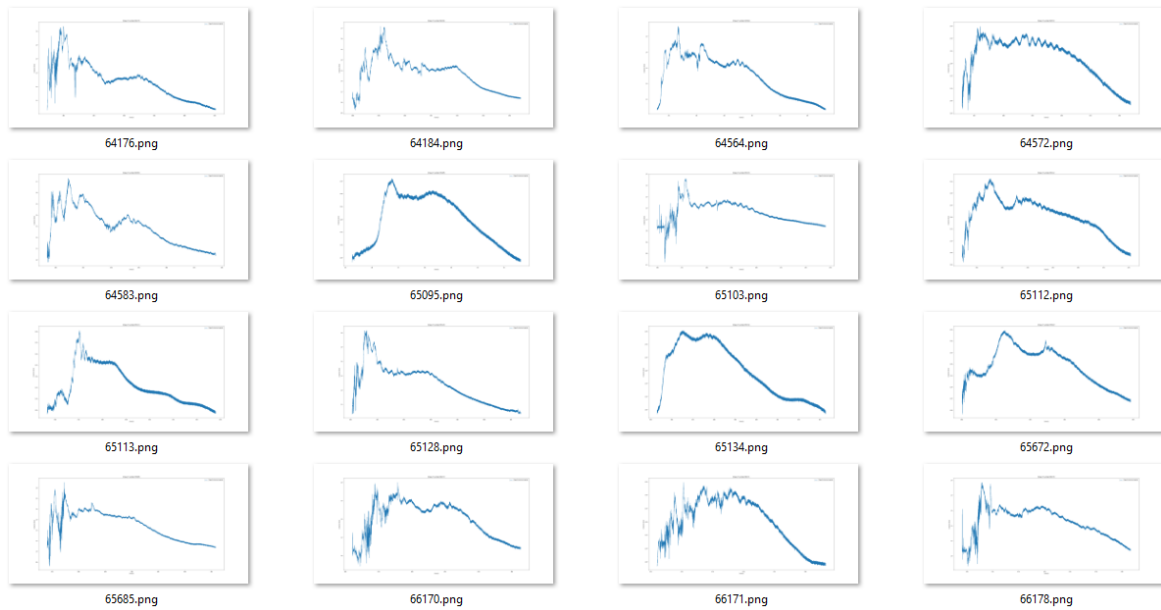


FIGURE 4.73 Echantillons de signaux d'impact du cluster 4 de K-Means.

Les figures 4.74, 4.75, 4.76, 4.77 et 4.78 montrent les signaux d'impact type par cluster. Ici aussi, en comparant ces signaux types avec les signaux types obtenus grâce au *Hierarchical Clustering*, nous pouvons facilement voir des similitudes entre les clusters présentés juste avant.

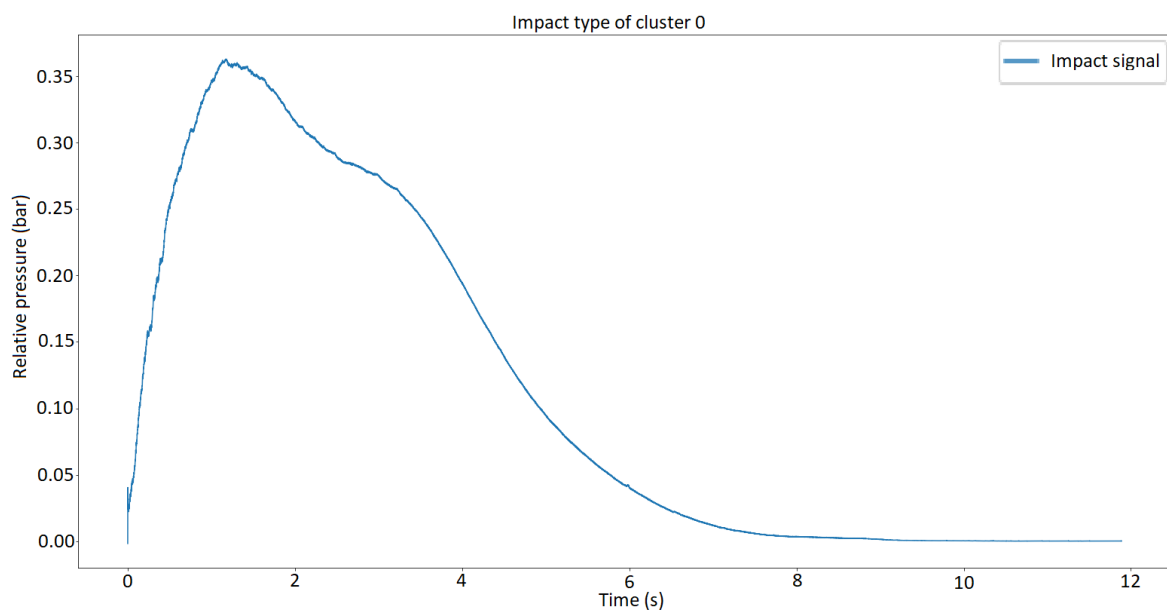


FIGURE 4.74 Signal d'impact type moyenné pour le cluster 0 obtenu avec l'algorithme K-Means.

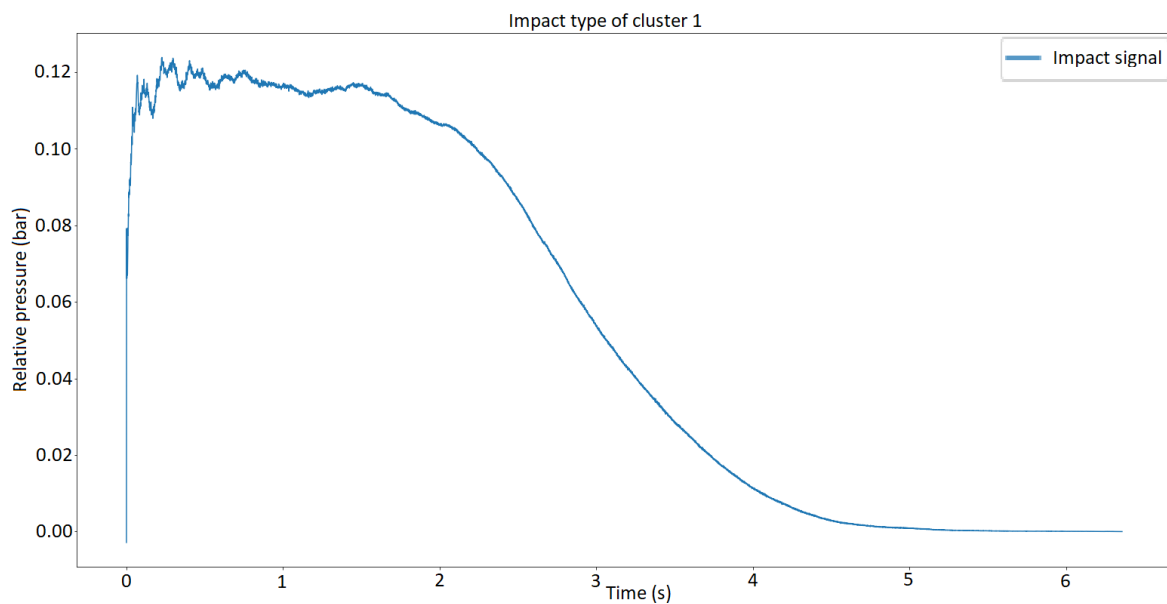


FIGURE 4.75 Signal d'impact type moyenné pour le cluster 1 obtenu avec l'algorithme K-Means.

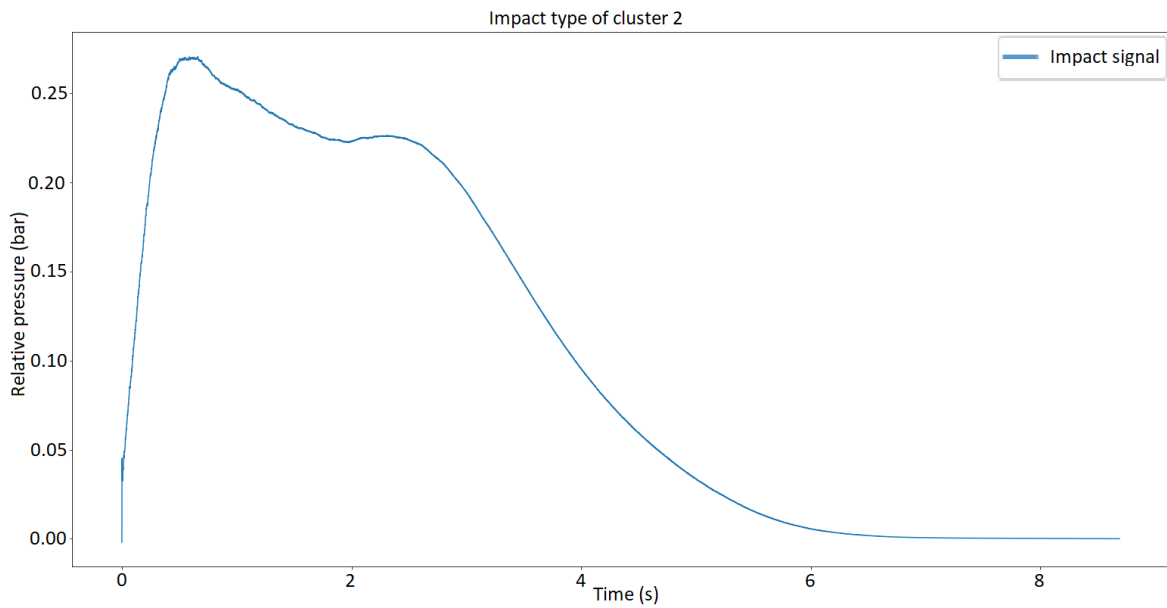


FIGURE 4.76 Signal d'impact type moyenné pour le cluster 2 obtenu avec l'algorithme K-Means.

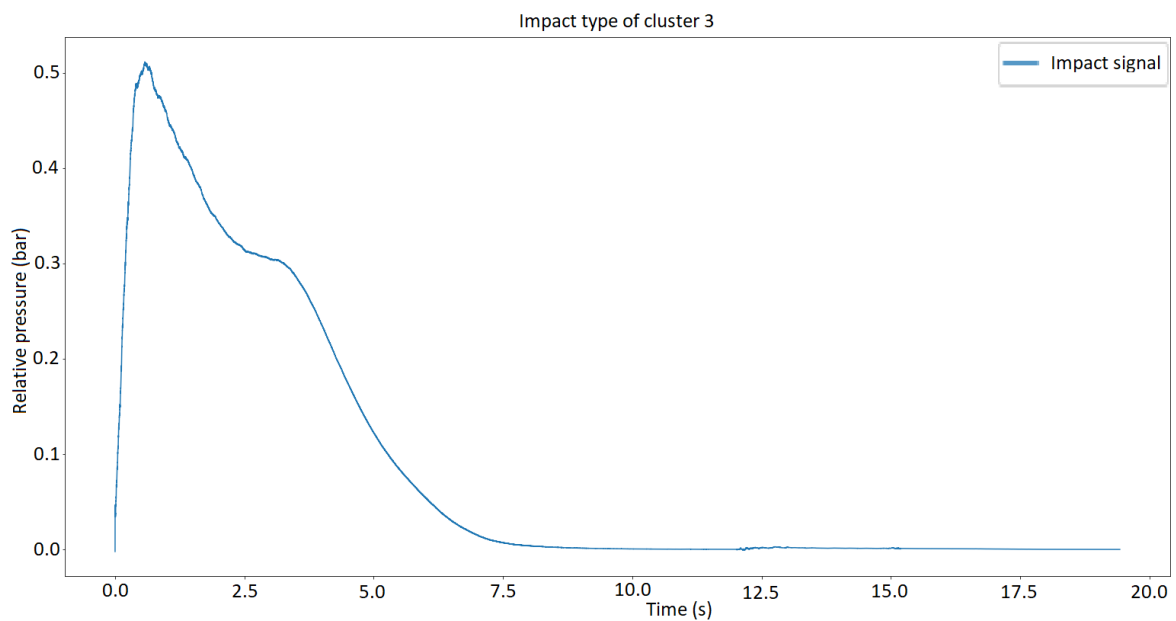


FIGURE 4.77 Signal d'impact type moyenné pour le cluster 3 obtenu avec l'algorithme K-Means.

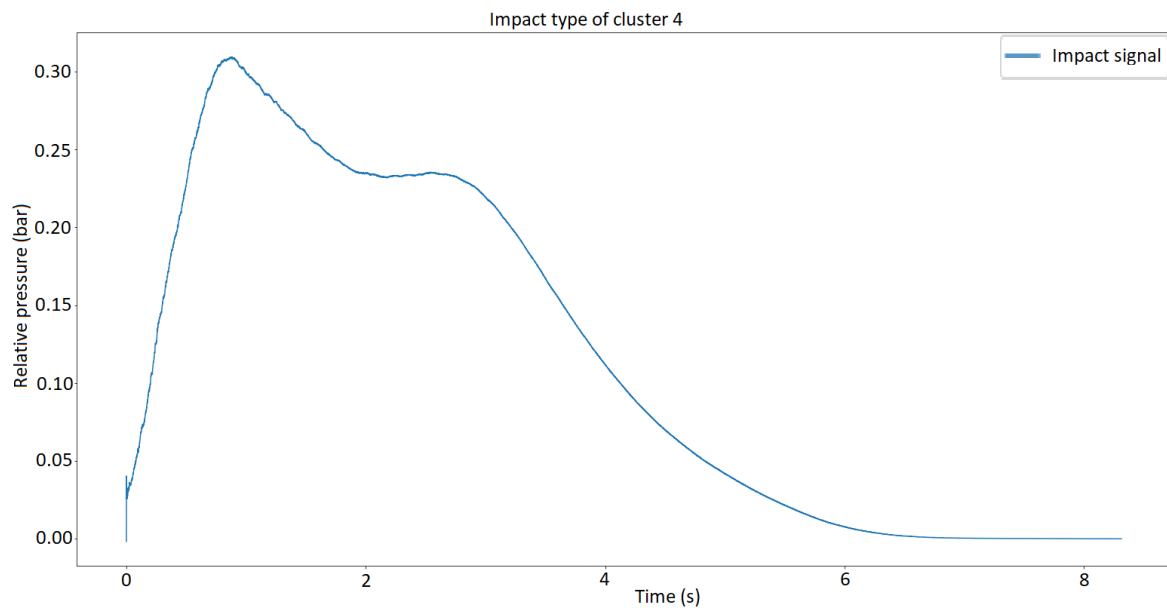


FIGURE 4.78 Signal d'impact type moyenné pour le cluster 4 obtenu avec l'algorithme K-Means.

De la même manière que dans la section 4.8.9, nous avons réalisé le tableau 4.79 permettant de visualiser le nombre et le pourcentage d'impacts en commun entre chaque cluster obtenu avec *Hierarchical Clustering* (HC) et K-Means (KM). Sur ce tableau nous pouvons bien observer qu'il y a beaucoup de similitudes entre ces deux regroupements, bien qu'ils ne soient pas réellement équivalents. Un point très positif que nous pouvons observer est qu'il y a beaucoup de lignes où le nombre d'impacts en commun est égale à 0. Ceci démontre que le clustering des deux algorithmes est proche.

HC Clusters id	Nombre d'impacts total	KM Clusters id	Nombre d'impacts total	Nombre d'impacts en commun	Pourcentage en commun par rapport au cluster HC	Pourcentage en commun par rapport au cluster KM
0	1612	0	235	0	0,00	0,00
		1	415	216	13,40	52,05
		2	1263	1127	69,91	89,23
		3	449	18	1,12	4,01
		4	750	251	15,57	33,47
1	687	0	235	2	0,29	0,85
		1	415	0	0,00	0,00
		2	1263	136	19,80	10,77
		3	449	396	57,64	88,20
		4	750	153	22,27	20,40
2	158	0	235	152	96,20	64,68
		1	415	0	0,00	0,00
		2	1263	0	0,00	0,00
		3	449	6	3,80	1,34
		4	750	0	0,00	0,00
3	199	0	235	0	0,00	0,00
		1	415	199	100,00	47,95
		2	1263	0	0,00	0,00
		3	449	0	0,00	0,00
		4	750	0	0,00	0,00
4	456	0	235	81	17,76	34,47
		1	415	0	0,00	0,00
		2	1263	0	0,00	0,00
		3	449	29	6,36	6,46
		4	750	346	75,88	46,13

FIGURE 4.79 Comparaison des impacts en communs entre *Hierarchical Clustering* et K-Means.

4.8.11 Conclusion

Nous pouvons établir plusieurs conclusions de cette étude. La première est que l'algorithme DBSCAN n'a pas fonctionné pour notre jeu de données. En effet, cet algorithme ne nous a pas permis de bien regrouper les impacts en différents clusters.

D'après l'étude menée grâce aux trois autres algorithmes, le clustering obtenu avec SOM ne semble pas pertinent, aussi bien quand nous l'analysons seul, que lorsque nous le comparons aux deux autres regroupements.

En effet, les résultats obtenus avec *Hierarchical Clustering* et K-Means semblent plus pertinents, et sont plus proches l'un de l'autre. Cependant, il y a quand même des différences entre ces résultats. Il est complexe de déterminer lequel des deux est à sélectionner.

Ensuite, nous pouvons constater que malgré l'aide d'algorithmes de clustering, le résultat final est fortement sujet à interprétation et est de ce fait discutable. Seul l'algorithme K-Means donne un résultat complètement objectif.

Un défaut de ce regroupement est qu'aucun cluster n'est constitué seulement d'impacts ayant un pic très court (dont le temps de montée et le temps de descente sont rapides). En effet, ce type d'impact est présent mais n'est pas isolé dans un cluster à part, quel que soit l'algorithme.

Il ne faut pas oublier que le clustering identifie les similitudes entre les objets, qu'il regroupe en fonction de leurs caractéristiques communes et qui les différencie des autres groupes d'objets. Le résultat d'un clustering n'est donc pas à prendre comme une classification pure qui utilise des classes prédéfinies et permet donc d'être plus précis.

Conclusion et perspectives

Aujourd'hui, grâce au progrès technologique, nous avons pu mettre en place une station de mesure autonome, qui enregistre la pression des impacts des vagues sur une structure côtière placée en plein milieu de l'eau. Cette station permet d'obtenir des données de pression d'impacts en continu, dans toutes les configurations météorologiques possibles. Grâce à cela nous avons obtenu un volume important et varié de données, qui est assez vaste pour étudier la pression des impacts de vagues avec une approche informatique.

Afin d'utiliser ces données, nous avons mis en place un algorithme de prétraitement, permettant de supprimer le bruit qui pourrait s'être introduit dans les signaux. Cet algorithme repose principalement sur l'application d'un filtre non-linéaire puis d'un filtre linéaire. En vue d'étudier les impacts de vagues, il a fallu trouver un moyen efficace d'extraire les impacts des signaux. Dans ce but, nous avons mis en place un algorithme de détection automatique des impacts. Etant un point central de notre travail, nous avons testé plusieurs approches pour cet algorithme, de manière à optimiser les résultats de celui-ci. Pour pouvoir utiliser au mieux les impacts, nous avons déterminé plusieurs critères qui sont calculés automatiquement pour chaque signal de pression d'impact, permettant de caractériser chaque impact. Ceci nous a permis de réaliser deux bases de données qui nous ont été utiles pour nos travaux, mais qui sont aussi très utiles pour les experts du domaine. La première base de données est une base de données par impact, et la deuxième base de données est une base de données par enregistrement de 10 minutes.

Ensuite, nous avons travaillé avec la covariance, permettant de connaître le décalage dans le temps d'un impact de vague entre deux capteurs différents. En utilisant la covariance sur un signal comprenant plusieurs impacts, nous pouvons également extraire la période des vagues grâce à la transformée de Fourier.

Nous avons par la suite essayé de représenter l'impact d'une vague sur les différents capteurs dans une vidéo. L'objectif est de pouvoir observer, de façon plus visuelle que sur un signal, comment la vague percute la digue à l'emplacement des différents capteurs et

la pression qu'elle engendre à ces différents points. Dans cet objectif, nous avons réalisé deux vidéos, la première montrant l'impact d'une vague sur les 11 capteurs placés sur un axe vertical (1D), et la deuxième montrant l'impact d'une vague en 2D sur les 16 capteurs.

Nous avons également travaillé sur la prévision de puissance d'impact de vagues. Pour cela, nous avons utilisé les données des bases de données que nous avons générées, ainsi que des données provenant de différentes stations enregistrant les paramètres environnementaux locaux. Pour ce faire, nous avons utilisé un algorithme d'apprentissage supervisé, un réseau de neurones. Malheureusement, les résultats obtenus n'ont pas été convaincants et ont montré qu'il nous manquait de l'information pour réaliser une telle prédiction.

C'est pourquoi nous avons commencé à étudier les types d'impacts. Nous aurions souhaité pouvoir classifier les impacts provenant de la station de la digue de l'Artha grâce aux caractéristiques disponibles sur les types d'impacts mis en avant par les expériences en laboratoire, mais la différence d'échelle a rendu cela impossible. C'est pour cette raison que nous avons utilisé des algorithmes d'apprentissage non-supervisé pour réaliser un *clustering* des impacts que nous avons enregistrés. Ce *clustering* nous a permis de mettre en avant 5 types d'impacts différents.

Pour continuer ces travaux, beaucoup de pistes sont à explorer. Pour essayer de prévoir la puissance d'impact des vagues sur la structure, il pourrait être plus judicieux d'utiliser les données environnementales de stations qui se trouvent en plein milieu de l'océan. En effet, les vagues sont créées au large et donc utiliser les données environnementales de l'endroit où elles sont générées semble plus pertinent que d'utiliser les données locales, c'est-à-dire de saint-Jean-de-Luz. Mais cela pose une problématique majeure, qui est de réussir à déterminer avec assez de précision quelle vague générée au large a percuté la digue de l'Artha à un moment donné. Malgré les formules existantes sur la propagation des vagues, cela reste complexe car il est nécessaire d'avoir la bathymétrie du fond marin. De plus, le coût en termes de temps de calcul devient élevé.

Concernant le *clustering* des impacts, il serait tout à fait envisageable de réaliser un *clustering* non-supervisé en utilisant les réseaux de neurones à convolution. Plutôt que de se baser sur des critères calculés que nous avons déterminés, une telle approche permettrait de réaliser un regroupement des impacts en comparant directement l'allure des signaux.

Bibliographie

- [1] WH Munk. Origin and generation of waves. Technical report, Scripps Institution of Oceanography La Jolla Calif, 1951.
- [2] A Toffoli and EM Bitner-Gregersen. Types of ocean surface waves, wave classification. *Encyclopedia of Maritime and Offshore Engineering*, pages 1–8, 2017.
- [3] National Oceanic and Atmospheric Administration. Why does the ocean have waves ? <https://oceanservice.noaa.gov/facts/wavesinocean.html>, 2017.
- [4] LH Holthuijsen. *Waves in oceanic and coastal waters*. Cambridge university press, 2010.
- [5] IR Young. *Wind generated ocean waves*, volume 2. Elsevier, 1999.
- [6] AC Duxbury. *Fundamentals of oceanography*. USA : McGRAW-HILL, 2002.
- [7] J Guo. Simple and explicit solution of wave dispersion equation. *Coastal Engineering*, 45(2) :71–74, 2002.
- [8] PR Pinet. *Essential Invitation to Oceanography*. Jones & Bartlett Publishers, 2012.
- [9] CJ Galvin. Wave breaking in shallow water. *Waves on beaches and resulting sediment transport*, pages 413–456, 1972.
- [10] T Sunamura and SI Okazaki. Breaker types and wave reflection coefficient : Laboratory relationships. *Journal of coastal research*, pages 240–245, 1996.
- [11] K Mangor. Wave transformation. http://www.coastalwiki.org/wiki/Wave_transformation, 2020.
- [12] B Przedwojski. Bed topography and local scour in rivers with banks protected by groynes. *Journal of Hydraulic Research*, 33(2) :257–273, 1995.
- [13] D Isebe, F Bouchette, B Mohammadi, P Azerad, A Lambert, N Bujan, F Grasso, and H Michallet. Une nouvelle approche pour la protection des plages : applications à la plage du lido de sète. *Revue Paralia*, (14-16 octobre 2008, Sophia-Antipolis) :263–272, 2008.
- [14] JW Kamphuis. *Introduction to coastal engineering and management*, volume 48. World Scientific, 2020.

- [15] B Shipman and T Stojanovic. Facts, fictions, and failures of integrated coastal zone management in europe. *Coastal management*, 35(2-3) :375–398, 2007.
- [16] ISA Kabir Sadeghi and HF Albab. Classification of seawalls and their failure : An overview. 2018.
- [17] A Mertzanis, F Marabini, MG Angeli, K Mertzanis, F Pontoni, and P Gasparetto. Coastal zone management and special coastal protection works in vulnerable lagoon ecosystems : The case of “venice lagoon”(italy) and “north amvrakikos gulf lagoons complex”(greece).
- [18] C Ciria. Cetmef (2007)-the rock manual. the use of rock in hydraulic engineering. *Publicação*, (C683), 2007.
- [19] MJ Henriques, A Fonseca, D Roque, JN Lima, and J Marnoto. Assessing the quality of an uav-based orthomosaic and surface model of a breakwater. In *FIG Congress*, 2014.
- [20] JA Morales, M Cantano, A Rodríguez-Ramírez, and RM Banda. Mapping geomorphology and active processes on the coast of huelva (southwestern spain). *Journal of Coastal Research*, pages 89–99, 2006.
- [21] T Mikami, T Shibayama, M Esteban, and R Matsumaru. Field survey of the 2011 tohoku earthquake and tsunami in miyagi and fukushima prefectures. *Coastal Engineering Journal*, 54(1) :1250011–1, 2012.
- [22] S Takahashi. Design of vertical breakwaters. *PHRI reference document nr. 34*, 2002.
- [23] PAD Bird, AR Crawford, PJ Hewson, and GN Bullock. An instrument for field measurement of wave impact pressures and seawater aeration. *Coastal engineering*, 35(1-2) :103–122, 1998.
- [24] British Standards Institution. Code of practice for maritime structures, bs 6349, part 1, general criteria, 1984.
- [25] PA Blackmore and PJ Hewson. Experiments on full-scale wave impact pressures. *Coastal Engineering*, 8(4) :331–346, 1984.
- [26] Shore Protection Manual. Department of the army. *US Army Corps of Engineers*, 1984.
- [27] RA Bagnold. Interim report on wave-pressure research. *Journal of the Institution of Civil Engineers*, 12(7) :202–226, 1939.
- [28] G Müller and TJT Whittaker. Evaluation of design wave impact pressures. *Journal of waterway, port, coastal, and ocean engineering*, 122(1) :55–58, 1996.
- [29] A de Rouville. Etat actuel des études internationales sur les efforts dus aux lames. In *Annales Ponts et Chaussées*, volume 108, pages 5–113, 1938.
- [30] G Bullock, C Obhrai, G Müller, G Wolters, H Peregrine, and H Bredmose. Field and laboratory measurement of wave impacts. In *Coastal structures 2003*, pages 343–355. 2004.

- [31] GN Bullock, AR Crawford, PJ Hewson, MJA Walkden, and PAD Bird. The influence of air and scale on wave impact pressures. *Coastal Engineering*, 42(4) :291–312, 2001.
- [32] B Hofland, M Kaminski, and G Wolters. Large scale wave impacts on a vertical wall. *Coastal Engineering Proceedings*, 1(32) :15, 2010.
- [33] H Oumeraci, A Kortenhuis, Wi Allsop, M de Groot, R Crouch, H Vrijling, and H Voortman. *Probabilistic design tools for vertical breakwaters*. CRC Press, 2001.
- [34] P Hull and G Müller. An investigation of breaker heights, shapes and pressures. *Ocean Engineering*, 29(1) :59–79, 2002.
- [35] GN Bullock, C Obhrai, DH Peregrine, and H Bredmose. Violent breaking wave impacts. part 1 : Results from large-scale regular wave tests on vertical and sloping walls. *Coastal Engineering*, 54(8) :602–617, 2007.
- [36] DH Peregrine. Water-wave impact on walls. *Annual review of fluid mechanics*, 35(1) :23–43, 2003.
- [37] ML Kaminski and H Bogaert. Full scale sloshing impact tests. In *The Nineteenth International Offshore and Polar Engineering Conference*. International Society of Offshore and Polar Engineers, 2009.
- [38] JL de Lagrange. Mémoire sur la théorie du mouvement des fluides. *Oeuvres complètes*, 695, 1781.
- [39] DJ Hand. Principles of data mining. *Drug safety*, 30(7) :621–622, 2007.
- [40] L Rokach and O Maimon. *Data mining with decision trees : theory and applications*, volume 69. World scientific, 2008.
- [41] U Fayyad, G Piatetsky-Shapiro, and P Smyth. From data mining to knowledge discovery in databases. *AI magazine*, 17(3) :37–37, 1996.
- [42] JH Friedman. Data mining and statistics : What’s the connection ? *Computing Science and Statistics*, 29(1) :3–9, 1998.
- [43] O Maimon and L Rokach. *Data mining and knowledge discovery handbook*. 2005.
- [44] YA LeCun, L Bottou, GB Orr, and KR Müller. Efficient backprop. In *Neural networks : Tricks of the trade*, pages 9–48. Springer, 1998.
- [45] MT Hagan, HB Demuth, MH Beale, and O De Jesús. *Neural network design*, volume 20. Pws Pub. Boston, 1996.
- [46] Kajal Kiran Gulhare, SP Shukla, and LK Sharma. Deep neural network classification method to alzheimer’s disease detection. *Int J Adv Res Comput Sci Softw Eng*, 7(6) :1–4, 2017.
- [47] Mohsen Afana, Jomana Ahmed, Bayan Harb, Bassem S Abu-Nasser, and Samy S Abu-Naser. Artificial neural network for forecasting car mileage per gallon in the city. 2018.

- [48] M Mouadil. Introduction au deep learning : les réseaux de neurones. <https://meritis.fr/data/deep-learning/>, 2018.
- [49] DE Rumelhart and JL McClelland. Parallel distributed processing : explorations in the microstructure of cognition. volume 1. foundations. 1986.
- [50] X Glorot, A Bordes, and Y Bengio. Deep sparse rectifier neural networks. In *Proceedings of the fourteenth international conference on artificial intelligence and statistics*, pages 315–323, 2011.
- [51] B Schölkopf, AJ Smola, and F Bach. *Learning with kernels : support vector machines, regularization, optimization, and beyond*. MIT press, 2002.
- [52] MTR Ullas, M Begom, A Ahmed, and R Sultana. *A machine learning approach to detect depression and anxiety using supervised learning*. PhD thesis, Brac University, 2019.
- [53] MA Ebrahimi, Mohammad-Hadi Khoshtaghaza, Saeid Minaei, and Bahareh Jamshidi. Vision-based pest detection based on svm classification method. *Computers and Electronics in Agriculture*, 137 :52–58, 2017.
- [54] J MacQueen. Some methods for classification and analysis of multivariate observations. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, volume 1, pages 281–297. Oakland, CA, USA, 1967.
- [55] C Piech. K means. <https://stanford.edu/~cpiech/cs221/handouts/kmeans.html>.
- [56] Abdallah Moubayed, Mohammadnoor Injadat, Abdallah Shami, and Hanan Lutfiyya. Student engagement level in an e-learning environment : Clustering using k-means. *American Journal of Distance Education*, 34(2) :137–156, 2020.
- [57] A Vellido, K Gibert, C Angulo, and JDM Guerrero. *Advances in Self-Organizing Maps, Learning Vector Quantization, Clustering and Data Visualization : Proceedings of the 13th International Workshop, WSOM+ 2019, Barcelona, Spain, June 26-28, 2019*, volume 976. Springer, 2019.
- [58] T Kohonen. Exploration of very large databases by self-organizing maps. In *Proceedings of international conference on neural networks (icnn'97)*, volume 1, pages PL1–PL6. IEEE, 1997.
- [59] Wikipedia. Self-organizing map. https://en.wikipedia.org/wiki/Self-organizing_map.
- [60] Dubravko Miljković. Brief review of self-organizing maps. In *2017 40th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, pages 1061–1066. IEEE, 2017.
- [61] M Ester, HP Kriegel, J Sander, and X Xu. A density-based algorithm for discovering clusters in large spatial databases with noise. In *Kdd*, volume 96, pages 226–231, 1996.
- [62] K Khan, SU Rehman, K Aziz, S Fong, and S Sarasvady. Dbscan : Past, present and future. In *The fifth international conference on the applications of digital information and web technologies (ICADIWT 2014)*, pages 232–238. IEEE, 2014.

- [63] Wikipedia. Bdscan. <https://en.wikipedia.org/wiki/DBSCAN>.
- [64] Mingrui Zhang. Use density-based spatial clustering of applications with noise (dbscan) algorithm to identify galaxy cluster members. In *IOP Conference Series : Earth and Environmental Science*, volume 252, page 042033. IOP Publishing, 2019.
- [65] GJ Szekely and ML Rizzo. Hierarchical clustering via joint between-within distances : Extending ward's minimum variance method. *Journal of classification*, 22(2), 2005.
- [66] Samanta Rosati, Valentina Agostini, Marco Knaflitz, and Gabriella Balestra. Muscle activation patterns during gait : A hierarchical clustering analysis. *Biomedical Signal Processing and Control*, 31 :463–469, 2017.
- [67] J Gareth, W Daniela, H Trevor, and T Robert. *An introduction to statistical learning : with applications in R*. Springer, 2013.
- [68] M Kuhn and K Johnson. *Applied predictive modeling*, volume 26. Springer, 2013.
- [69] R Khandelwal. K fold and other cross-validation techniques. *Data Driven Investor*, [Online]. Available : <https://medium.com/datadriveninvestor/k-fold-and-other-cross-validation-techniques-6c03a2563f1e>, 2018.
- [70] S Raschka. Model evaluation, model selection, and algorithm selection in machine learning. *arXiv preprint arXiv :1811.12808*, 2018.
- [71] C Lugni, M Brocchini, and OM Faltinsen. Wave impact loads : The role of the flip-through. *Physics of fluids*, 18(12) :122101, 2006.
- [72] YM Scolan. Some aspects of the flip-through phenomenon : A numerical study based on the desingularized technique. *Journal of fluids and structures*, 26(6) :918–953, 2010.
- [73] A Rahali. Analyse des données issues des enregistrements des impacts réels des vagues sur la digue d'Artha. liuppa, université de pau et des pays l'adour. 2017.
- [74] B Larroque, P Arnould, F Luthon, PA Poncet, A Rahali, and S Abadie. In-situ measurements of wave impact pressure on a composite breakwater : preliminary results. *Journal of Coastal Research*, pages 1086–1090, 2018.
- [75] F Luthon, D D'Amico, and B Larroque. Embedded sensors in coastal structures for on-site wave impact pressure monitoring. *IEEE Access*, 6 :56135–56141, 2018.
- [76] Les ouvrages de protection. <https://www.saint-jean-de-luz.com/fr/decouvrir-saint-jean-de-luz/histoire-et-patrimoine/les-ouvrages-de-protection/>.
- [77] La tempete de decembre 1951 a st jean de luz en labourd au pays basque. <https://paysbasqueavant.blogspot.com/2017/03/la-tempete-de-decembre-1951-st-jean-de.html>.
- [78] Bloc de 50 tonnes déposé sur une digue à saint-jean-de-luz : "j'ai eu une chance incroyable". <https://www.sudouest.fr/2017/03/15/bloc-de-50-tonnes-depose-sur-une-digue-a-saint-jean-de-luz-j-ai-eu-une-chance-incroyable-3279061.php>.

- [79] Functionality and technology of pressure sensors. https://www.baumer.com/dk/en/service-support/know-how/function-principle/functionality-and-technology-of-pressure-sensors/a/Know-how_Function_Pressure-sensors.
- [80] National instruments. <https://www.ni.com>.
- [81] CE Shannon. Communication in the presence of noise. *Proceedings of the IRE*, 37(1) :10–21, 1949.
- [82] Conseil départemental 64. <https://www.le64.fr/>.
- [83] PA Poncet, B Liqueur, B Larroque, D d'Amico, and S Abadie. In-situ measurements of wave impact pressure on a composite breakwater submitted to high wave energy : data analysis and influence of the environmental variables. *Coastal Engineering*, 2020.
- [84] rp5.am. <http://rp5.am/>.
- [85] Data.shom.fr. <https://data.shom.fr/>.
- [86] Candhis. <http://candhis.cetmef.developpement-durable.gouv.fr/>.
- [87] P Du, WA Kibbe, and SM Lin. Improved peak detection in mass spectrum by incorporating continuous wavelet transform-based pattern matching. *Bioinformatics*, 22(17) :2059–2065, 2006.
- [88] MJ Cooker and DH Peregrine. A model for breaking wave impact pressures. In *Coastal Engineering 1990*, pages 1473–1486. 1991.
- [89] Y Goda. New wave pressure formulae for composite breakwaters. In *Coastal Engineering 1974*, pages 1702–1720. 1975.
- [90] D d'Amico, B Larroque, F Luthon, PA Poncet, and S Abadie. Continuous measurement and automatic processing of in-situ wave impact pressure data. *Journal of Coastal Research*, 95(sp1) :214–219, 2020.
- [91] J Vesanto, J Himberg, E Alhoniemi, and J Parhankangas. Som toolbox for matlab 5. Technical report, Citeseer, 2000.
- [92] D d'Amico, B Larroque, F Luthon, PA Poncet, and S Abadie. Continuous measurement and automatic processing of in-situ wave impact pressure data. *International Coastal Symposium (ICS 2020)*, Seville(Spain) :20–23th April, 2020.
- [93] D D'Amico, B Larroque, and F Luthon. Mesure de la pression d'impact des vagues in-situ. *Journée plénière de la fédération de recherche MIRA*, Pau(France) :20 Février, 2020.
- [94] D D'Amico, B Larroque, and F Luthon. Mesure de la pression d'impact des vagues in-situ. *Journée des doctorants organisée par la CAPB (Communauté d'Agglomération du Pays Basque)*, Bayonne(France) :18 Décembre, 2018.

Publications et conférences

Publications revues internationales

Publiées

1. D d'Amico, B Larroque, F Luthon, PA Poncet, and S Abadie. Continuous measurement and automatic processing of in-situ wave impact pressure data. *Journal of Coastal Research*, 95(sp1) :214–219, 2020
2. F Luthon, D D'Amico, and B Larroque. Embedded sensors in coastal structures for on-site wave impact pressure monitoring. *IEEE Access*, 6 :56135–56141, 2018

Soumise

3. PA Poncet, B Liquez, B Larroque, D d'Amico, and S Abadie. In-situ measurements of wave impact pressure on a composite breakwater submitted to high wave energy : data analysis and influence of the environmental variables. *Coastal Engineering*, 2020

Conférences

Internationales

1. D d'Amico, B Larroque, F Luthon, PA Poncet, and S Abadie. Continuous measurement and automatic processing of in-situ wave impact pressure data. *International Coastal Symposium (ICS 2020)*, Seville(Spain) :20–23th April, 2020

Nationale

2. D D'Amico, B Larroque, and F Luthon. Mesure de la pression d'impact des vagues in-situ. *Journée plénière de la fédération de recherche MIRA*, Pau(France) :20 Février, 2020

Autre

3. D D'Amico, B Larroque, and F Luthon. Mesure de la pression d'impact des vagues in-situ. *Journée des doctorants organisée par la CAPB (Communauté d'Agglomération du Pays Basque)*, Bayonne(France) :18 Décembre, 2018

Annexe A

Extraits des tableaux de données

A.1 Tableaux des données environnementales

A.1.1 Extrait des données de la station météo de Socoa

Heure locale à Socoa	T	Po	P	Pa	U	DD	Ff	ff10	ff3	N	WW	W1
19.03.2019 04:00	9.0	766.9	770.9			85 Le vent qui souffle du nord		6	8	12 100%.	Pluie, sans congélation, inte Pluie.	
13.03.2019 13:00	11.7	768.3	772.3			72 Le vent qui souffle de l'ouest-nord-ouest		10	13	90 ou plus, mais pas 100°	'Etat du ciel inchang'e dans l' Pluie.	
06.03.2019 07:00	18.4	746.5	750.3	-1.5		44 Le vent qui souffle du sud-sud-est		8	14	23 60%.		
28.01.2019 07:00	11.0	753.2	757.1	-2.7		89 Le vent qui souffle de l'ouest		19	29	29 100%.	Pluie, sans congélation, inte Pluie.	
27.01.2019 07:00	9.4	754.6	758.5	-0.3		79 Le vent qui souffle de l'ouest-sud-ouest		16	24	25 100%.	Pluie, sans congélation, inte Pluie.	
23.01.2019 22:00	8.8	753.1	757.0			81 Le vent qui souffle du nord-nord-ouest		14	24	25 100%.	Pluie, sans congélation, inte Pluie.	
23.01.2019 07:00	11.9	746.3	750.1	-3.0		90 Le vent qui souffle de l'ouest		22	28	28 100%.	Pluie, sans congélation, inte Pluie.	
21.12.2018 22:00	15.3	766.8	770.7			58 Le vent qui souffle du sud		7	10	50%.		
29.11.2018 07:00	15.7	756.5	760.3	-0.6		31 Le vent qui souffle du sud		9	16	17 90 ou plus, mais pas 100°		
10.11.2018 04:00	16.8	751.5	755.3			55 Le vent qui souffle du sud-sud-ouest		8	14	17 70 – 80%.	Dans l'ensemble, nuages se Pluie.	
07.11.2018 07:00	15.5	750.7	754.6	0.7		55 Le vent qui souffle du sud		11	18	21 70 – 80%.	'Etat du ciel inchang'e dans l' Pluie.	
05.11.2018 07:00	18.4	745.9	749.7	-0.1		44 Le vent qui souffle du sud-sud-est		10	19	19 100%.		
18.09.2018 11:00	23.9	759.4	763.2			75 Le vent qui souffle de l'ouest		3	3	70 – 80%.		
18.09.2018 05:00	21.0	758.2	762.0			78 Le vent qui souffle de l'ouest-sud-ouest		4	7	8 Pas de nuages.		
17.09.2018 11:00	24.6	758.3	762.1			70 Le vent qui souffle de l'est-sud-est		3	8	8 Pas de nuages.		
17.09.2018 05:00	19.8	760.3	764.2			89 Le vent qui souffle de l'est		3	4	4 70 – 80%.		
17.09.2018 02:00	20.1	761.7	765.5			88 Le vent qui souffle du sud		3	3	3 90 ou plus, mais pas 100°		
16.09.2018 11:00	22.8	761.8	765.5			77 Le vent qui souffle de l'ouest		5	7	40%.		
16.09.2018 02:00	17.3	760.9	764.8			90 Le vent qui souffle du sud		2	3	3 Pas de nuages.		
15.09.2018 23:00	19.0	761.4	765.2			85 Le vent qui souffle du sud		2	3	3 Pas de nuages.		
15.09.2018 20:00	21.8	761.2	765.0			77 Le vent qui souffle du nord-nord-est		6	8	40%.		

FIGURE A.1 Début d'un extrait de données de la station météo de Socoa.

W1	W2	Tn	Tx	Cl	Nh	H	Cm	Ch	VV	Td	RRR	RR	E	Tg	E'	sss
Pluie.	Pluie.			Stratus fracti 100%.		600-1000			10.0	6.6	2.0		3	Surface du sol humide.		
Pluie.	Nuages couvrant plus de la moitié du côté Cumulus et S 90 ou plus, r 600-1000						Pas d'AltoCumulus, d'Altos		12.0	6.8	8.0		6	Surface du sol mouillée (eau stagnante en mares, petites ou gran		
		17.2	19.0	Pas de Strato 60%.		2500 ou plus	AltoCumulus	Pas de Cirrus 25.0	5.9	5.9		Pas de préc	12	S'èche (sans fissure et sans pouss'ère ni sable sec en quant'è a		
Pluie.	Pluie.	8.0	11.7	Stratus fracti 100%.		300-600			6.0	9.3	11.0		12	Surface du sol mouillée (eau stagnante en mares, petites ou gran		
Pluie.	Nuages couv 6.8		11.2	Stratus fracti 100%.		600-1000			12.0	5.9	2.0		12	Surface du sol mouillée (eau stagnante en mares, petites ou gran		
Pluie.	Pluie.			Stratus fracti 100%.		300-600			6.0	5.7	1.0		3	Surface du sol mouillée (eau stagnante en mares, petites ou gran		
Pluie.	Pluie.	6.5	12.2	Stratus fracti 100%.		300-600			8.0	10.3	24.0		12	Surface du sol mouillée (eau stagnante en mares, petites ou gran		
				Pas de Strato 20-30%.			AltoCumulus	Cirrostratus 1	18.0	7.1		Pas de préc	3	S'èche (sans fissure et sans pouss'ère ni sable sec en quant'è a		
		14.3	16.6	Stratus nebuli 40%.		1500-2000	AltoCumulus	translucidus & 20.0	20.0	-1.4	0.8		12	S'èche (sans fissure et sans pouss'ère ni sable sec en quant'è a		
Pluie.	Nuages couvrant plus de la moitié du côté Stratocumuli 70-80%.						Pas d'AltoCumulus, d'Altostratus, ni de N 7.7			6.5	0.8		3	Surface du sol humide.		
Pluie.	Nuages couv 13.4		16.1	Stratus fracti 40%.		1000-1500	AltoCumulus	translucidus & 20.0	20.0	5.9		Pas de préc	12	S'èche (sans fissure et sans pouss'ère ni sable sec en quant'è a		
		13.9	19.1	Pas de Strato 100%.		2000-2500	Altostratus opacus ou Nim	18.0	18.0	6.5	0.8		12	S'èche (sans fissure et sans pouss'ère ni sable sec en quant'è a		
				Pas de Strato 70-80%.			AltoCumulus	Pas de Cirrus 12.0	12.0	19.2		Pas de préc	3	S'èche (sans fissure et sans pouss'ère ni sable sec en quant'è a		
									20.0	17.0		Pas de préc	3	S'èche (sans fissure et sans pouss'ère ni sable sec en quant'è a		
									15.0	18.8		Pas de préc	3	S'èche (sans fissure et sans pouss'ère ni sable sec en quant'è a		
				Stratocumuli 70-80%.		600-1000	Pas d'AltoCuir	Pas de Cirrus 12.0	12.0	17.9		Pas de préc	6	S'èche (sans fissure et sans pouss'ère ni sable sec en quant'è a		
				Stratocumuli 90 ou plus, r 600-1000					12.0	18.0		Pas de préc	3	S'èche (sans fissure et sans pouss'ère ni sable sec en quant'è a		
				Pas de Strato 10% ou moins, mais pas 0			AltoCumulus	Cirrus spissal 12.0	12.0	18.6		Pas de préc	6	S'èche (sans fissure et sans pouss'ère ni sable sec en quant'è a		
									12.0	15.6		Pas de préc	3	S'èche (sans fissure et sans pouss'ère ni sable sec en quant'è a		
									20.0	16.4		Pas de préc	12	S'èche (sans fissure et sans pouss'ère ni sable sec en quant'è a		
		14.1	24.4						20.0	17.6		Pas de préc				

FIGURE A.2 Suite d'un extrait de données de la station météo de Socoa.

A.1.2 Extrait des données du marégraphe de Socoa

```
# Date;Valeur;Source
01/01/2019 00:00:00;3.217;2
01/01/2019 00:00:00;3.217;3
01/01/2019 00:00:27;3.2175;1
01/01/2019 00:01:27;3.218;1
01/01/2019 00:02:27;3.2175;1
01/01/2019 00:03:27;3.2173;1
01/01/2019 00:04:27;3.2181;1
01/01/2019 00:05:27;3.2155;1
01/01/2019 00:06:27;3.215;1
01/01/2019 00:07:27;3.2145;1
01/01/2019 00:08:27;3.212;1
01/01/2019 00:09:27;3.2093;1
01/01/2019 00:10:00;3.208;3
01/01/2019 00:10:00;3.208;2
01/01/2019 00:10:27;3.2072;1
01/01/2019 00:11:27;3.2047;1
01/01/2019 00:12:27;3.202;1
01/01/2019 00:13:27;3.1998;1
01/01/2019 00:14:27;3.1951;1
01/01/2019 00:15:27;3.1935;1
01/01/2019 00:16:27;3.192;1
01/01/2019 00:17:27;3.19;1
01/01/2019 00:18:27;3.1875;1
01/01/2019 00:19:27;3.185;1
01/01/2019 00:20:00;3.184;2
01/01/2019 00:20:00;3.184;3
```

FIGURE A.3 Extrait de données provenant du marégraphe de Socoa.

A.1.3 Extrait des données du houlographe

dateheure	h13d	h110d	hmaxd	hsigmad	hrmsd	h2percentd	th13d	th110d	tavd	thmaxd	tmaxd	etamax	etamin
01/01/2018 00:00	3.68	4.41	5.61	3.89	2.63	4.80	11.0	11.3	8.7	11.6	17.3	3.19	2.88
01/01/2018 00:30	3.45	4.20	5.55	3.78	2.47	4.79	10.5	11.1	7.9	14.2	17.5	3.06	2.77
01/01/2018 01:00	3.70	4.57	5.39	4.14	2.70	4.76	10.4	10.8	8.1	12.2	15.9	3.20	3.01
01/01/2018 01:30	3.57	4.42	6.50	4.27	2.57	4.72	9.9	10.5	7.5	10.6	16.9	3.41	3.49
01/01/2018 02:00	3.71	4.58	5.84	4.13	2.70	5.39	10.1	10.8	8.1	10.1	16.2	3.67	2.93
01/01/2018 02:30	3.68	4.72	6.22	3.95	2.58	5.21	10.8	11.6	7.9	11.9	17.5	3.53	3.33
01/01/2018 03:00	3.48	4.36	6.83	3.91	2.48	4.80	10.4	11.6	7.7	10.6	16.7	4.22	3.47
01/01/2018 03:30	3.70	4.67	6.50	3.99	2.59	5.07	10.8	11.5	8.0	12.8	18.3	3.36	3.16
01/01/2018 04:00	3.83	4.82	5.47	4.10	2.76	5.18	11.3	11.1	8.8	11.1	17.4	3.32	2.96
01/01/2018 04:30	3.96	4.90	6.11	4.26	2.84	5.41	11.7	12.5	9.1	10.0	18.0	3.59	3.12
01/01/2018 05:00	3.93	4.91	7.73	4.28	2.81	4.96	11.7	12.4	9.0	11.9	16.2	4.69	3.92
01/01/2018 05:30	4.30	5.39	6.57	4.39	2.98	6.07	12.5	12.0	9.5	11.0	17.5	3.74	3.47
01/01/2018 06:00	4.02	4.86	5.53	4.26	2.91	5.39	12.7	11.9	10.3	13.2	17.6	3.53	2.78
01/01/2018 06:30	3.98	4.79	6.48	4.23	2.84	5.28	12.6	13.6	10.1	13.4	18.6	3.82	2.96
01/01/2018 07:00	4.09	5.25	8.74	4.28	2.81	6.22	12.8	12.9	9.6	13.0	17.2	4.53	4.21
01/01/2018 07:30	3.92	4.75	6.79	4.11	2.86	5.03	13.0	12.6	10.9	13.5	18.0	3.43	3.36
01/01/2018 08:00	3.35	3.95	4.46	3.51	2.39	4.36	12.2	12.0	9.8	12.0	18.1	2.67	2.38
01/01/2018 08:30	3.79	4.86	6.07	4.04	2.69	5.48	12.8	13.2	10.2	12.6	18.1	3.07	3.02
01/01/2018 09:00	4.05	5.15	6.33	4.33	2.83	5.92	13.4	13.9	10.2	14.1	18.1	3.77	3.33

FIGURE A.4 Début d'un extrait de données du houlographe de Saint-Jean-de-Luz.

sz13d	szmaxd	tszmaxd	nbrevag	skew	kurt	rhh	hm0	tp	t02	te	eps2	kapa	thetap	thetam	sigmap	sigmam
0.046	0.144	1.50	184	0.092	2.790	0.19	3.82	13.5	8.3	10.8	0.48	0.27	309.7	306.7	35.5	31.0
0.057	0.218	1.70	199	0.088	2.797	0.19	3.80	15.1	7.8	10.5	0.51	0.23	304.8	306.5	38.2	32.4
0.049	0.105	3.70	193	-0.030	2.708	0.04	4.09	14.6	8.0	10.6	0.49	0.22	304.3	304.1	35.5	30.9
0.063	0.257	1.00	200	-0.023	2.791	0.13	4.05	14.7	7.8	10.6	0.51	0.22	306.7	304.9	34.3	31.8
0.065	0.743	0.90	187	0.162	2.904	0.15	4.15	14.1	7.9	10.5	0.49	0.22	307.9	303.7	34.9	32.2
0.055	0.178	1.20	198	0.039	3.180	0.18	3.88	14.1	7.9	10.8	0.53	0.28	315.5	308.3	28.9	30.9
0.050	0.129	2.50	205	-0.008	3.246	0.08	3.79	14.0	7.8	10.7	0.54	0.31	312.8	308.3	33.1	33.0
0.054	0.199	1.50	196	0.149	3.103	0.30	4.00	14.1	8.0	10.8	0.51	0.30	321.4	312.1	32.2	30.1
0.049	0.150	1.80	180	0.138	2.781	0.13	4.06	14.0	8.3	10.9	0.50	0.35	312.2	311.1	31.0	29.8
0.048	0.261	1.70	174	0.140	2.888	0.28	4.24	13.6	8.8	11.3	0.49	0.35	316.7	313.7	25.6	26.9
0.043	0.188	1.20	176	0.207	3.198	0.21	4.22	14.1	8.8	11.2	0.47	0.33	316.1	314.7	32.0	27.3
0.045	0.155	2.40	168	-0.033	3.014	0.18	4.48	14.0	9.3	12.0	0.49	0.48	313.3	313.8	30.2	28.4
0.043	0.459	1.00	154	0.271	2.880	0.23	4.20	14.0	9.4	11.9	0.47	0.48	316.8	318.0	29.5	27.6
0.038	0.244	1.20	156	0.163	2.837	0.34	4.18	14.0	9.7	12.1	0.47	0.50	317.5	317.0	27.9	28.1
0.036	0.073	2.70	165	0.198	3.769	0.41	4.06	14.6	9.8	12.3	0.46	0.48	314.4	314.9	30.3	28.6
0.029	0.052	4.80	146	0.222	2.897	0.39	4.18	14.0	10.0	12.3	0.44	0.43	314.6	314.9	29.0	26.9
0.029	0.057	3.30	162	0.127	2.685	0.15	3.47	14.1	9.3	11.9	0.48	0.41	309.5	313.7	32.6	30.8
0.031	0.082	2.50	155	0.162	2.939	0.45	3.94	14.0	10.0	12.4	0.45	0.47	311.4	314.0	33.6	30.8
0.032	0.070	2.00	155	0.080	3.033	0.52	4.22	14.1	10.3	12.6	0.44	0.48	315.0	314.7	30.5	29.1

FIGURE A.5 Suite d'un extrait de données du houlgraphe de Saint-Jean-de-Luz.

A.2 Bases de données réalisées

A.2.1 Base de données d'impacts

FileName	SensorId	Uniqueld	Startindex	StartTime	Endindex	EndTime	Maxindex	MaxValue	RelativeMaxValue	RiseTimeMax	DropTimeMax	Integrate	Duration
20181021-020001-10min-23063imp-rate21.raww	0	555	3276380	319,9589844	3320973	324,31377	3397803	1,219	0,164	2,079589844	2,252246094	0,282413525	4,35302734
20181021-020001-10min-23063imp-rate21.raww	0	556	3405159	332,5390586	3441149	336,049707	3419236	1,214	0,159	1,34453125	2,096289063	0,306905225	3,47688547
20181021-020001-10min-23063imp-rate21.raww	0	557	3538462	345,5529297	3563898	348,036914	3544106	1,148	0,093	0,549804688	1,92900306	0,132725049	2,48183594
20181021-020001-10min-23063imp-rate21.raww	0	558	3837530	374,7587891	3854937	376,458691	3844796	1,148	0,093	0,708398438	0,988476563	0,088277246	1,69873047
20181021-020001-10min-23063imp-rate21.raww	0	559	5095162	497,5744141	5117610	499,766602	5108311	1,156	0,101	1,280859375	0,9046875	0,113677295	2,18886719
20181021-020001-10min-23063imp-rate21.raww	0	560	5489916	531,1441406	5464307	533,62373	5455710	1,181	0,126	1,638867188	0,833886719	0,183734229	2,47675781
20181021-020001-10min-23063imp-rate21.raww	0	561	5567919	543,7420898	5606576	547,517188	5584798	1,203	0,148	1,612695313	2,071289063	0,287720459	3,73076172
20181021-020001-10min-23063imp-rate21.raww	0	562	5753016	561,8179688	5774444	563,910547	5760849	1,441	0,086	0,763476563	1,325976563	0,103145703	2,09121094
20181021-020001-10min-23063imp-rate21.raww	0	563	5868400	573,0859375	5911250	577,270508	5885393	1,213	0,158	1,645410156	2,48828125	0,325040283	4,17236328
20181021-020001-10min-23063imp-rate21.raww	0	564	6005410	586,4658203	6039034	589,749414	6009028	1,178	0,123	0,352050781	2,908300781	0,208646289	3,271875
20181021-020001-10min-23063imp-rate21.raww	1	565	2964637	289,515332	2980729	291,086816	2965123	1,151	0,075	0,046386719	1,522070313	0,061951953	1,57021484
20181021-020001-10min-23063imp-rate21.raww	1	566	3283534	320,6576172	3299635	322,22998	3291559	1,145	0,069	0,782421875	0,787402344	0,064815186	1,57119141
20181021-020001-10min-23063imp-rate21.raww	1	567	3404917	332,5114258	3425993	334,569629	3417799	1,149	0,073	1,25625	0,797949219	0,077537988	2,05644531
20181021-020001-10min-23063imp-rate21.raww	1	568	5442553	531,4993164	5457024	532,9125	5452922	1,145	0,069	1,010546875	0,39921875	0,051525293	1,41181641
20181021-020001-10min-23063imp-rate21.raww	1	569	5566093	543,5637695	5587614	545,66543	5566135	1,197	0,121	0,002929688	2,088378906	0,084791504	2,09257813
20181021-020001-10min-23063imp-rate21.raww	1	570	5869185	573,1625977	5890859	575,279199	5869224	1,27	0,194	0,002441406	2,040136719	0,11752793	2,04394531
20181021-020001-10min-23063imp-rate21.raww	1	571	6006440	586,5664063	6023901	588,271582	6008289	1,178	0,102	0,179785156	1,52109375	0,062031494	1,70234375
20181021-020001-10min-23063imp-rate21.raww	1	572	2965060	289,5566406	2979948	291,010547	2975681	1,132	0,055	1,036035156	0,415429688	0,0444092725	1,45283203
20181021-020001-10min-23063imp-rate21.raww	2	573	3284391	320,7413086	3298420	322,111328	3293266	1,136	0,059	0,865234375	0,502148438	0,045498975	1,36875
20181021-020001-10min-23063imp-rate21.raww	2	574	3405082	332,5275391	3424501	334,423926	3417630	1,133	0,056	1,23242188	0,669140625	0,051475439	1,89414063
20181021-020001-10min-23063imp-rate21.raww	2	575	5440476	531,2964844	5456057	532,818066	5453956	1,136	0,059	1,315429688	0,203613281	0,042951514	1,52050781
20181021-020001-10min-23063imp-rate21.raww	2	576	5570351	543,9795898	5587187	545,62373	5583004	1,144	0,067	1,233886719	0,406933594	0,055997266	1,64287109
20181021-020001-10min-23063imp-rate21.raww	2	577	5869436	573,1871094	5889670	575,163086	5884511	1,155	0,078	1,470117188	0,502148438	0,079253613	1,97441406
20181021-020001-10min-23063imp-rate21.raww	2	578	6010947	587,006543	6023465	588,229004	6019114	1,13	0,053	0,796386719	0,423339844	0,02957002	1,22158203
20181021-020001-10min-23063imp-rate21.raww	3	579	3287423	321,0374023	3292717	321,993848	3292161	1,117	0,058	0,461621094	0,492480469	0,032425146	0,95537109
20181021-020001-10min-23063imp-rate21.raww	3	580	5570776	544,0210938	5586074	545,515039	5581147	1,112	0,053	1,01171875	0,479980469	0,040712549	1,38287109
20181021-020001-10min-23063imp-rate21.raww	3	581	5873772	573,6105469	5887487	574,949992	5884483	1,137	0,078	1,044628906	0,291503906	0,048461719	1,338163959
20181021-020001-10min-23063imp-rate21.raww	4	582	3286237	320,921582	3292661	321,92002	3292593	1,122	0,051	0,619443394	0,376367188	0,026107031	0,99716797
20181021-020001-10min-23063imp-rate21.raww	4	583	5874215	573,6538086	5886541	574,85752	5883223	1,135	0,064	0,878222656	0,322460938	0,0383537178	1,20244141
20181021-020001-10min-23063imp-rate21.raww	5	584	5875326	573,7623047	5885980	574,802734	5882763	1,124	0,059	0,725	0,312695313	0,027849902	1,03925781

FIGURE A.6 Début de l'extrait de la base de données des impacts

PressureImpulse	ImpulseDuration	StaticPressureBeforeImpact	AssociatedImpacts	AssociationImpactId	LinearForceMax	Error
0.278939355	4.159179688		1.053276923 {1: 566, 2: 573, 3: 579, 4: 582}	555	22,1025	FAUX
0.266239893	2.6890625		1.05237536 {1: 567, 2: 574}	556	23,3825	FAUX
0.06135459	1.099609375		1.052497688 {}	557	12,2175	FAUX
0.082501807	1.416796875		1.052695857 {}	558	8,21	FAUX
0.113681738	2.189550781		1.052780699 {}	559	12,47	FAUX
0.183759619	2.479296875		1.052341215 {1: 568, 2: 575}	560	19,52	FAUX
0.275644189	3.225390625		1.052374632 {1: 569, 2: 576, 3: 580}	561	24,5425	FAUX
0.088973682	1.526953125		1.053048682 {}	562	8,26	FAUX
0.29618418	3.290820313		1.052007791 {1: 570, 2: 577, 3: 581, 4: 583, 5: 584, 6: 585, 11: 586, 12: 587}	563	27,4775	FAUX
0.056055078	0.704101563		1.052537654 {1: 571, 2: 578}	564	21,695	FAUX
0.004103516	0.092773438		1.073219637 {0: 554, 2: 572}	554		FAUX
0.064773242	1.56484375		1.073291053 {0: 555, 2: 573, 3: 579, 4: 582}	555		FAUX
0.077542529	2.057324219		1.073643668 {0: 556, 2: 574}	556		FAUX
0.051527637	1.412304688		1.073647994 {0: 560, 2: 575}	560		FAUX
0.000417969	0.005859375		1.073361848 {0: 561, 2: 576, 3: 580}	561		FAUX
0.000688379	0.004882813		1.073647082 {0: 563, 2: 577, 3: 581, 4: 583, 5: 584, 6: 585, 11: 586, 12: 587}	563		FAUX
0.018917432	0.359570313		1.0734410573 {0: 564, 2: 578}	564		FAUX
0.04409458	1.453320313		1.073691172 {0: 554, 1: 565}	554		FAUX
0.045500537	1.369140625		1.073773238 {0: 555, 1: 566, 3: 579, 4: 582}	555		FAUX
0.051478467	1.895019531		1.073985252 {0: 556, 1: 567}	556		FAUX
0.042953857	1.521191406		1.074027377 {0: 560, 1: 568}	560		FAUX
0.056	1.643554688		1.07434562 {0: 561, 1: 569, 3: 580}	561		FAUX
0.079257471	1.975195313		1.07407272 {0: 563, 1: 570, 3: 581, 4: 583, 5: 584, 6: 585, 11: 586, 12: 587}	563		FAUX
0.029572217	1.222167969		1.074913099 {0: 564, 1: 571}	564		FAUX
0.032079932	0.923242188		1.055955187 {0: 555, 1: 566, 2: 573, 4: 582}	555		FAUX
0.040714404	1.493359375		1.056385383 {0: 561, 1: 569, 2: 576}	561		FAUX
0.048465088	1.338769531		1.056189296 {0: 563, 1: 570, 2: 577, 4: 583, 5: 584, 6: 585, 11: 586, 12: 587}	563		FAUX
0.026109326	0.997753906		1.067758684 {0: 555, 1: 566, 2: 573, 3: 579}	555		FAUX
0.038359473	1.202929688		1.068080828 {0: 563, 1: 570, 2: 577, 3: 581, 5: 584, 6: 585, 11: 586, 12: 587}	563		FAUX
0.027852246	1.03984375		1.06223622 {0: 563, 1: 570, 2: 577, 3: 581, 4: 583, 6: 585, 11: 586, 12: 587}	563		FAUX

FIGURE A.7 Suite de l'extrait de la base de données des impacts

Annexe B

Tables de données liées au réseaux de neurones de la section 4.7

B.1 Format des données utilisé pour le réseau de neurones

P1/10	h110d	h110d	h110d	WA_N	WA_NNE	WA_NNE	WA_NNE	WA_ENE	WA_E	WA_ESE	WA_SE	WA_SSE	WA_S	WA_SSW	WA_SW	WA_WSW	WA_W	WA_WNW	WA_NW	WA_NNW	W_N	
1,1402	1,88	9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	3,96	11,1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
0,5406	1,91	9,6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
0,76566667	4,17	15,2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
0,83866667	5,39	11,1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
0,9215	6,59	14,1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
0	4,38	15,3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
0	3,94	15,3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
0	4	16,6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
0,36166667	2,2	7,1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
0,4105	1,27	7,4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
0	4,5	15,9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
0,717	4,22	12,8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
0,6006	2,57	10,5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
0,48233333	2,33	11,9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
0	4,28	15,7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
0,67475	2,75	13,3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
0,7116	3,92	12,6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
0	4,26	15,9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
0,75266667	4,93	13	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
0	3,57	13,1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
0,76975	2,94	13,3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
0,43025	3,29	8,7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
0	4,03	12,7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
0,47725	2,77	13,9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
0	3,16	11,5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
1,17333333	4,76	13,5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
0,51325	2,92	7,8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
0,60375	3,75	13,6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1

FIGURE B.1 Début d'un extrait de données formatées.

Wl_N	Wl_NNE	Wl_NE	Wl_ENE	Wl_E	Wl_ESE	Wl_SE	Wl_SSE	Wl_S	Wl_SSW	Wl_SW	Wl_WSW	Wl_W	Wl_WNW	Wl_NW	Wl_NNW	Ef	Valeur
0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	3	4,388
0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	14	3,147
0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	3	4,275
0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	2	4,291
0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	7	2,066
0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	12	3,34
0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	7	1,51
0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	6	1,363
0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	5	0,937
0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	12	4,272
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	3,461
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	8	1,231
0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	5	3,659
0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	6	3,224
0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	2	4,119
0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	2	2,921
0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	4,079
0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	4	4,799
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	7	2,032
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	3,489
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	1,01
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	4,492
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	3,831
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	2,123
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	3,453
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	7	3,184
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	2,877
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	11	3,283
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	3,869

FIGURE B.2 Suite d'un extrait de données formatées.

B.2 Tableau des données de validation du réseau de neurones

La tableau B.1 montre les 770 valeurs P1/10 de test que le modèle entraîné a essayé de déterminer grâce aux données environnementales. Ce tableau contient également les valeurs P1/10 calculées par le modèle, ainsi que la différence absolue entre la vraie valeur du P1/10 et la valeur prédit.

N°	P110 réel	P110 prédit	Diff absolue
0	0.246000	0.192198	0.053802
1	0.000000	-0.000803	0.000803
2	0.168000	-0.000803	0.168803
3	0.000000	-0.000803	0.000803
4	0.000000	-0.000803	0.000803
5	0.000000	-0.000803	0.000803
6	0.114333	-0.000803	0.115136
7	0.268000	0.238264	0.029736
8	0.000000	-0.000803	0.000803
9	0.410500	-0.000803	0.411303
10	0.000000	-0.000803	0.000803
11	0.497500	0.756080	0.258580
12	0.000000	0.187759	0.187759
13	0.000000	-0.000803	0.000803
14	0.000000	-0.000803	0.000803
15	0.311000	0.276782	0.034218
16	0.211667	0.219398	0.007731
17	0.136000	0.270271	0.134271
18	0.000000	-0.000803	0.000803
19	0.154000	-0.000803	0.154803
20	0.000000	-0.000803	0.000803
21	0.000000	-0.000803	0.000803
22	0.195500	0.270901	0.075401
23	0.252500	0.327321	0.074821
24	0.000000	-0.000803	0.000803
25	0.259000	-0.000803	0.259803
26	0.000000	-0.000803	0.000803

27	0.230333	0.170054	0.060279
28	0.135000	0.193709	0.058709
29	0.000000	-0.000803	0.000803
30	0.000000	-0.000803	0.000803
31	0.000000	-0.000803	0.000803
32	0.152333	0.033532	0.118801
33	0.000000	-0.000803	0.000803
34	0.000000	-0.000803	0.000803
35	0.146333	0.189236	0.042903
36	0.000000	-0.000803	0.000803
37	0.000000	-0.000803	0.000803
38	0.250400	0.175765	0.074635
39	0.000000	-0.000803	0.000803
40	0.000000	-0.000803	0.000803
41	0.000000	-0.000803	0.000803
42	0.000000	-0.000803	0.000803
43	0.133000	0.173517	0.040517
44	0.000000	-0.000803	0.000803
45	0.282000	0.178529	0.103471
46	0.105000	-0.000803	0.105803
47	0.000000	0.187436	0.187436
48	0.000000	-0.000803	0.000803
49	0.447500	0.242361	0.205139
50	0.000000	-0.000803	0.000803
51	0.282500	0.175765	0.106735
52	0.000000	-0.000803	0.000803
53	0.000000	-0.000803	0.000803
54	0.000000	-0.000803	0.000803
55	0.431000	0.153341	0.277659
56	0.000000	-0.000803	0.000803
57	0.000000	-0.000803	0.000803
58	0.455667	0.404732	0.050935
59	0.000000	-0.000803	0.000803
60	0.273000	0.354410	0.081410
61	0.000000	-0.000803	0.000803
62	0.000000	-0.000803	0.000803

63	0.000000	-0.000803	0.000803
64	0.000000	0.200970	0.200970
65	0.000000	-0.000803	0.000803
66	0.000000	-0.000803	0.000803
67	0.205500	-0.000803	0.206303
68	0.186000	-0.000803	0.186803
69	0.000000	-0.000803	0.000803
70	0.000000	-0.000803	0.000803
71	0.788000	-0.000803	0.788803
72	0.000000	0.190680	0.190680
73	0.306000	0.179023	0.126977
74	0.277000	-0.000803	0.277803
75	0.000000	0.175765	0.175765
76	0.171000	0.205786	0.034786
77	0.297000	-0.000803	0.297803
78	0.324667	0.240677	0.083990
79	0.000000	-0.000803	0.000803
80	0.000000	-0.000803	0.000803
81	0.000000	-0.000803	0.000803
82	0.000000	-0.000803	0.000803
83	0.000000	0.122845	0.122845
84	0.000000	-0.000803	0.000803
85	0.000000	-0.000803	0.000803
86	0.000000	-0.000803	0.000803
87	0.229400	0.293819	0.064419
88	0.000000	-0.000803	0.000803
89	0.139000	-0.000803	0.139803
90	0.000000	-0.000803	0.000803
91	0.430250	0.185771	0.244479
92	0.000000	-0.000803	0.000803
93	0.000000	-0.000803	0.000803
94	0.282000	0.175765	0.106235
95	0.275000	0.221682	0.053318
96	0.000000	-0.000803	0.000803
97	0.000000	0.410079	0.410079
98	0.000000	-0.000803	0.000803

99	0.222250	-0.000803	0.223053
100	0.000000	-0.000803	0.000803
101	0.398667	0.242816	0.155851
102	0.187750	-0.000803	0.188553
103	0.207000	-0.000803	0.207803
104	0.000000	0.025673	0.025673
105	0.000000	-0.000803	0.000803
106	0.000000	-0.000803	0.000803
107	0.000000	-0.000803	0.000803
108	0.370250	0.357658	0.012592
109	0.000000	0.212132	0.212132
110	0.000000	0.124532	0.124532
111	0.000000	-0.000803	0.000803
112	0.000000	0.058000	0.058000
113	0.000000	-0.000803	0.000803
114	0.000000	-0.000803	0.000803
115	0.000000	-0.000803	0.000803
116	0.000000	-0.000803	0.000803
117	0.000000	-0.000803	0.000803
118	0.000000	-0.000803	0.000803
119	0.000000	-0.000803	0.000803
120	0.000000	-0.000803	0.000803
121	0.000000	-0.000803	0.000803
122	0.000000	-0.000803	0.000803
123	0.000000	-0.000803	0.000803
124	0.000000	-0.000803	0.000803
125	0.000000	-0.000803	0.000803
126	0.241000	-0.000803	0.241803
127	0.000000	-0.000803	0.000803
128	0.000000	-0.000803	0.000803
129	0.000000	-0.000803	0.000803
130	0.000000	-0.000803	0.000803
131	0.000000	-0.000803	0.000803
132	0.000000	-0.000803	0.000803
133	0.000000	0.175765	0.175765
134	0.000000	0.184411	0.184411

135	0.215000	-0.000803	0.215803
136	0.000000	-0.000803	0.000803
137	0.000000	-0.000803	0.000803
138	0.179000	0.458190	0.279190
139	0.000000	-0.000803	0.000803
140	0.000000	-0.000803	0.000803
141	0.416500	0.463826	0.047326
142	0.000000	-0.000803	0.000803
143	0.000000	-0.000803	0.000803
144	0.000000	-0.000803	0.000803
145	0.000000	-0.000803	0.000803
146	0.000000	-0.000803	0.000803
147	0.000000	-0.000803	0.000803
148	0.323000	0.175765	0.147235
149	0.000000	-0.000803	0.000803
150	0.000000	-0.000803	0.000803
151	0.278000	-0.000371	0.278371
152	0.000000	-0.000803	0.000803
153	0.000000	-0.000803	0.000803
154	0.235000	-0.000803	0.235803
155	0.337500	0.562753	0.225253
156	0.000000	-0.000803	0.000803
157	0.000000	-0.000803	0.000803
158	0.000000	-0.000803	0.000803
159	0.118333	-0.000803	0.119136
160	0.000000	-0.000803	0.000803
161	0.000000	0.287678	0.287678
162	0.251333	0.188193	0.063140
163	0.516750	-0.000803	0.517553
164	0.000000	-0.000803	0.000803
165	0.000000	0.008738	0.008738
166	0.000000	0.185599	0.185599
167	0.000000	0.267681	0.267681
168	0.000000	-0.000803	0.000803
169	0.550250	0.459505	0.090745
170	0.623000	0.595769	0.027231

171	0.260000	0.332921	0.072921
172	0.266143	-0.000803	0.266946
173	0.525333	0.443467	0.081866
174	0.000000	-0.000803	0.000803
175	0.000000	-0.000803	0.000803
176	0.000000	-0.000803	0.000803
177	0.421500	0.303892	0.117608
178	0.000000	-0.000803	0.000803
179	0.000000	-0.000803	0.000803
180	0.000000	-0.000803	0.000803
181	0.000000	-0.000803	0.000803
182	0.000000	-0.000803	0.000803
183	0.259750	0.122498	0.137252
184	0.206667	-0.000803	0.207470
185	0.000000	-0.000803	0.000803
186	0.000000	-0.000803	0.000803
187	0.186500	0.169084	0.017416
188	0.000000	-0.000803	0.000803
189	0.000000	-0.000803	0.000803
190	0.241400	0.222485	0.018915
191	0.420000	0.343510	0.076490
192	0.000000	-0.000803	0.000803
193	0.502333	0.196791	0.305542
194	0.174333	-0.000803	0.175136
195	0.000000	-0.000803	0.000803
196	0.240000	-0.000803	0.240803
197	0.104000	-0.000803	0.104803
198	0.440000	0.490595	0.050595
199	0.000000	-0.000803	0.000803
200	0.000000	-0.000803	0.000803
201	0.000000	-0.000803	0.000803
202	0.305000	0.215422	0.089578
203	0.140250	0.183937	0.043687
204	0.000000	-0.000803	0.000803
205	0.000000	-0.000803	0.000803
206	0.101000	-0.000803	0.101803

207	0.199667	-0.000803	0.200470
208	0.000000	-0.000803	0.000803
209	0.678600	0.486063	0.192537
210	0.260000	0.175765	0.084235
211	0.270000	0.259655	0.010345
212	0.000000	-0.000803	0.000803
213	0.000000	-0.000803	0.000803
214	0.242000	0.175765	0.066235
215	0.000000	-0.000803	0.000803
216	0.234000	-0.000803	0.234803
217	0.000000	-0.000803	0.000803
218	0.000000	-0.000803	0.000803
219	0.184250	-0.000803	0.185053
220	0.198000	0.175765	0.022235
221	0.285500	0.297337	0.011837
222	0.389000	-0.000803	0.389803
223	0.000000	-0.000803	0.000803
224	0.310000	0.263875	0.046125
225	0.000000	-0.000803	0.000803
226	0.441000	-0.000803	0.441803
227	0.000000	-0.000803	0.000803
228	0.214400	0.282725	0.068325
229	0.000000	0.270206	0.270206
230	0.676000	0.323432	0.352568
231	0.000000	-0.000803	0.000803
232	0.721000	-0.000803	0.721803
233	0.100000	-0.000803	0.100803
234	0.224000	-0.000803	0.224803
235	0.000000	0.211268	0.211268
236	0.000000	-0.000803	0.000803
237	1.000670	0.822234	0.178436
238	0.000000	-0.000803	0.000803
239	0.000000	-0.000803	0.000803
240	0.000000	-0.000803	0.000803
241	0.000000	0.213429	0.213429
242	0.381000	0.133999	0.247001

243	0.000000	-0.000803	0.000803
244	0.000000	-0.000803	0.000803
245	0.000000	-0.000803	0.000803
246	0.137333	-0.000803	0.138136
247	0.526000	-0.000803	0.526803
248	0.000000	-0.000803	0.000803
249	0.000000	-0.000803	0.000803
250	0.000000	0.071822	0.071822
251	0.155000	-0.000803	0.155803
252	0.126000	-0.000803	0.126803
253	0.000000	-0.000803	0.000803
254	0.000000	-0.000803	0.000803
255	0.240333	0.306961	0.066628
256	0.000000	0.295291	0.295291
257	0.000000	-0.000803	0.000803
258	0.000000	-0.000803	0.000803
259	0.000000	0.175765	0.175765
260	0.000000	-0.000803	0.000803
261	0.118000	0.090556	0.027444
262	0.000000	-0.000803	0.000803
263	1.017500	0.683105	0.334395
264	0.244500	0.175765	0.068735
265	0.000000	-0.000803	0.000803
266	0.623750	0.608220	0.015530
267	0.293667	0.101329	0.192338
268	0.000000	0.249862	0.249862
269	0.000000	-0.000803	0.000803
270	0.000000	-0.000803	0.000803
271	0.000000	-0.000803	0.000803
272	0.134000	-0.000803	0.134803
273	0.339500	-0.000803	0.340303
274	0.000000	-0.000803	0.000803
275	0.207000	-0.000803	0.207803
276	0.000000	-0.000803	0.000803
277	0.000000	0.187744	0.187744
278	0.352333	0.403516	0.051183

279	0.000000	-0.000803	0.000803
280	0.112000	-0.000803	0.112803
281	0.000000	-0.000803	0.000803
282	0.176500	-0.000803	0.177303
283	0.000000	-0.000803	0.000803
284	0.153000	0.292662	0.139662
285	0.000000	-0.000803	0.000803
286	0.227000	0.339889	0.112889
287	0.000000	-0.000803	0.000803
288	0.243750	-0.000803	0.244553
289	0.111000	-0.000803	0.111803
290	0.000000	0.181723	0.181723
291	0.000000	-0.000803	0.000803
292	0.000000	-0.000803	0.000803
293	0.000000	-0.000803	0.000803
294	0.000000	-0.000803	0.000803
295	0.000000	-0.000803	0.000803
296	0.000000	-0.000803	0.000803
297	0.000000	0.187399	0.187399
298	0.000000	-0.000803	0.000803
299	0.000000	-0.000803	0.000803
300	0.000000	-0.000803	0.000803
301	0.000000	-0.000803	0.000803
302	0.000000	-0.000803	0.000803
303	0.162000	-0.000803	0.162803
304	0.529000	0.374037	0.154963
305	0.000000	-0.000803	0.000803
306	0.000000	-0.000803	0.000803
307	0.074000	-0.000803	0.074803
308	0.000000	-0.000803	0.000803
309	0.114250	0.175765	0.061515
310	0.328500	0.363629	0.035129
311	0.000000	-0.000803	0.000803
312	0.000000	-0.000803	0.000803
313	0.385000	0.015509	0.369491
314	0.000000	-0.000803	0.000803

315	0.000000	-0.000803	0.000803
316	0.000000	-0.000803	0.000803
317	0.225667	0.266841	0.041174
318	0.000000	-0.000803	0.000803
319	0.000000	-0.000803	0.000803
320	0.489000	-0.000803	0.489803
321	0.750500	0.424412	0.326088
322	0.000000	-0.000803	0.000803
323	0.000000	0.218060	0.218060
324	0.000000	-0.000803	0.000803
325	0.203667	0.155779	0.047888
326	0.000000	-0.000803	0.000803
327	0.000000	-0.000803	0.000803
328	0.212500	-0.000803	0.213303
329	0.000000	-0.000803	0.000803
330	0.000000	-0.000803	0.000803
331	0.300000	0.371101	0.071101
332	0.000000	-0.000803	0.000803
333	0.354000	0.288515	0.065485
334	0.110000	-0.000803	0.110803
335	0.000000	-0.000803	0.000803
336	0.000000	-0.000803	0.000803
337	0.499000	0.264055	0.234945
338	0.216000	0.214873	0.001127
339	0.000000	-0.000803	0.000803
340	0.000000	0.047118	0.047118
341	0.185500	-0.000803	0.186303
342	0.149333	-0.000803	0.150136
343	0.108333	-0.000803	0.109136
344	0.193500	0.209040	0.015540
345	0.574250	0.315179	0.259071
346	0.000000	-0.000803	0.000803
347	0.242500	-0.000803	0.243303
348	0.775000	0.186344	0.588656
349	0.692000	0.530937	0.161063
350	0.319000	0.479386	0.160386

351	0.000000	-0.000803	0.000803
352	0.495000	0.241794	0.253206
353	0.223667	0.285571	0.061904
354	0.000000	-0.000803	0.000803
355	0.000000	-0.000803	0.000803
356	0.000000	-0.000803	0.000803
357	0.000000	-0.000803	0.000803
358	0.117500	-0.000803	0.118303
359	0.133333	0.109240	0.024093
360	0.235500	0.199248	0.036252
361	0.281500	0.175486	0.106014
362	0.000000	-0.000803	0.000803
363	0.307000	-0.000803	0.307803
364	0.388000	0.426040	0.038040
365	0.000000	-0.000803	0.000803
366	0.000000	-0.000803	0.000803
367	0.000000	-0.000803	0.000803
368	0.000000	-0.000803	0.000803
369	0.000000	-0.000803	0.000803
370	0.000000	-0.000803	0.000803
371	0.000000	-0.000803	0.000803
372	0.000000	-0.000803	0.000803
373	0.000000	-0.000803	0.000803
374	0.000000	-0.000803	0.000803
375	0.000000	-0.000803	0.000803
376	0.000000	-0.000803	0.000803
377	0.000000	0.470434	0.470434
378	0.000000	-0.000803	0.000803
379	0.000000	-0.000803	0.000803
380	0.000000	0.342803	0.342803
381	0.000000	-0.000803	0.000803
382	0.000000	-0.000803	0.000803
383	0.532000	0.328282	0.203718
384	0.000000	-0.000803	0.000803
385	0.000000	-0.000803	0.000803
386	1.047000	0.865110	0.181890

387	0.000000	-0.000803	0.000803
388	0.000000	-0.000803	0.000803
389	0.361667	0.175765	0.185902
390	0.000000	-0.000803	0.000803
391	0.000000	-0.000803	0.000803
392	0.000000	-0.000803	0.000803
393	0.344250	0.345591	0.001341
394	0.000000	-0.000803	0.000803
395	0.000000	-0.000803	0.000803
396	0.000000	-0.000803	0.000803
397	0.124667	-0.000803	0.125470
398	0.000000	-0.000803	0.000803
399	0.431333	0.288007	0.143326
400	0.000000	-0.000803	0.000803
401	0.000000	0.253095	0.253095
402	0.000000	-0.000803	0.000803
403	0.000000	-0.000803	0.000803
404	0.000000	-0.000803	0.000803
405	0.000000	-0.000803	0.000803
406	0.000000	-0.000803	0.000803
407	0.000000	-0.000803	0.000803
408	0.233250	0.195610	0.037640
409	0.262500	0.188908	0.073592
410	0.000000	0.165653	0.165653
411	0.000000	-0.000803	0.000803
412	0.483500	0.562607	0.079107
413	0.000000	-0.000803	0.000803
414	0.000000	-0.000803	0.000803
415	0.000000	-0.000803	0.000803
416	0.000000	-0.000803	0.000803
417	0.276250	-0.000803	0.277053
418	0.000000	-0.000803	0.000803
419	0.000000	-0.000803	0.000803
420	0.099000	-0.000803	0.099803
421	0.000000	0.140656	0.140656
422	0.000000	-0.000803	0.000803

423	0.000000	-0.000803	0.000803
424	0.242000	-0.000803	0.242803
425	0.000000	-0.000803	0.000803
426	0.000000	-0.000803	0.000803
427	0.214833	0.315215	0.100382
428	0.000000	-0.000803	0.000803
429	0.000000	-0.000803	0.000803
430	0.330750	0.310496	0.020254
431	0.390667	0.340371	0.050296
432	0.127500	0.012536	0.114964
433	0.000000	-0.000803	0.000803
434	0.758400	0.608780	0.149620
435	0.000000	-0.000803	0.000803
436	0.206000	-0.000803	0.206803
437	0.221600	-0.000803	0.222403
438	0.121000	-0.000803	0.121803
439	0.000000	-0.000803	0.000803
440	0.000000	-0.000803	0.000803
441	0.000000	-0.000803	0.000803
442	0.544000	0.363164	0.180836
443	0.000000	-0.000803	0.000803
444	0.000000	0.147075	0.147075
445	0.297000	0.314797	0.017797
446	1.026000	0.925622	0.100378
447	0.252500	0.274507	0.022007
448	0.000000	-0.000803	0.000803
449	0.000000	0.103111	0.103111
450	0.000000	-0.000803	0.000803
451	0.134000	-0.000803	0.134803
452	0.000000	-0.000803	0.000803
453	0.684667	-0.000803	0.685470
454	0.000000	-0.000803	0.000803
455	0.000000	-0.000803	0.000803
456	0.096500	-0.000803	0.097303
457	0.000000	-0.000803	0.000803
458	0.379250	0.745667	0.366417

459	0.000000	-0.000803	0.000803
460	0.000000	-0.000803	0.000803
461	0.000000	-0.000803	0.000803
462	0.102000	-0.000803	0.102803
463	0.000000	-0.000803	0.000803
464	0.000000	0.191962	0.191962
465	0.000000	0.175765	0.175765
466	0.000000	-0.000803	0.000803
467	0.123000	0.212812	0.089812
468	0.000000	-0.000803	0.000803
469	0.000000	-0.000803	0.000803
470	0.119333	0.175765	0.056432
471	0.000000	-0.000803	0.000803
472	0.000000	-0.000803	0.000803
473	0.000000	-0.000803	0.000803
474	0.000000	-0.000803	0.000803
475	0.000000	-0.000803	0.000803
476	0.000000	-0.000803	0.000803
477	0.000000	-0.000803	0.000803
478	0.233000	0.274975	0.041975
479	0.112000	-0.000803	0.112803
480	0.105000	0.298953	0.193953
481	0.413000	0.343569	0.069431
482	0.456333	0.445103	0.011230
483	0.000000	-0.000803	0.000803
484	0.449667	0.382517	0.067150
485	0.000000	-0.000803	0.000803
486	0.000000	-0.000803	0.000803
487	0.000000	-0.000803	0.000803
488	0.280000	-0.000803	0.280803
489	0.000000	-0.000803	0.000803
490	0.129000	-0.000803	0.129803
491	0.213500	-0.000803	0.214303
492	0.000000	-0.000803	0.000803
493	0.250000	0.144507	0.105493
494	0.000000	-0.000803	0.000803

495	0.198667	-0.000803	0.199470
496	0.000000	-0.000803	0.000803
497	0.000000	-0.000803	0.000803
498	0.671000	0.304350	0.366650
499	0.000000	-0.000803	0.000803
500	0.104000	-0.000803	0.104803
501	0.000000	-0.000803	0.000803
502	0.000000	-0.000803	0.000803
503	0.000000	-0.000803	0.000803
504	0.386000	-0.000803	0.386803
505	0.000000	-0.000803	0.000803
506	0.000000	-0.000803	0.000803
507	0.519000	0.456436	0.062564
508	0.294750	0.297128	0.002378
509	0.166000	-0.000803	0.166803
510	0.509000	0.322524	0.186476
511	0.143000	-0.000803	0.143803
512	0.288333	-0.000803	0.289136
513	0.000000	-0.000803	0.000803
514	0.000000	-0.000803	0.000803
515	0.000000	-0.000803	0.000803
516	0.000000	-0.000803	0.000803
517	0.421250	0.429669	0.008419
518	0.000000	-0.000803	0.000803
519	0.000000	-0.000803	0.000803
520	0.000000	-0.000803	0.000803
521	0.000000	-0.000803	0.000803
522	0.202000	-0.000803	0.202803
523	0.000000	-0.000803	0.000803
524	0.167667	0.170935	0.003268
525	0.000000	-0.000803	0.000803
526	0.000000	-0.000803	0.000803
527	0.000000	-0.000803	0.000803
528	0.457500	-0.000803	0.458303
529	0.000000	-0.000803	0.000803
530	0.123500	-0.000803	0.124303

531	0.000000	-0.000803	0.000803
532	0.819333	0.196995	0.622338
533	0.000000	-0.000803	0.000803
534	0.104000	0.014371	0.089629
535	0.000000	-0.000803	0.000803
536	0.000000	-0.000803	0.000803
537	0.252500	0.064029	0.188471
538	0.000000	-0.000803	0.000803
539	0.599500	0.331676	0.267824
540	0.000000	-0.000803	0.000803
541	0.324750	0.187748	0.137002
542	0.000000	-0.000803	0.000803
543	0.000000	-0.000803	0.000803
544	0.341333	-0.000803	0.342136
545	1.355500	0.702998	0.652502
546	0.233500	0.273345	0.039845
547	0.000000	-0.000803	0.000803
548	0.000000	-0.000803	0.000803
549	0.000000	-0.000803	0.000803
550	0.151667	0.076682	0.074985
551	0.000000	-0.000803	0.000803
552	0.000000	-0.000803	0.000803
553	0.178250	0.391884	0.213634
554	0.000000	-0.000803	0.000803
555	0.000000	-0.000803	0.000803
556	0.000000	-0.000803	0.000803
557	0.000000	-0.000803	0.000803
558	0.101000	-0.000803	0.101803
559	0.000000	-0.000803	0.000803
560	0.000000	-0.000803	0.000803
561	0.101000	-0.000803	0.101803
562	0.000000	0.294593	0.294593
563	0.000000	-0.000803	0.000803
564	0.000000	-0.000803	0.000803
565	0.000000	-0.000803	0.000803
566	0.000000	-0.000803	0.000803

567	0.387000	0.140836	0.246164
568	0.000000	-0.000803	0.000803
569	0.000000	-0.000803	0.000803
570	0.000000	0.282364	0.282364
571	0.502000	0.265691	0.236309
572	0.100000	-0.000803	0.100803
573	0.000000	-0.000803	0.000803
574	0.291667	-0.000803	0.292470
575	0.000000	-0.000803	0.000803
576	0.000000	0.358490	0.358490
577	0.000000	-0.000803	0.000803
578	0.000000	0.187630	0.187630
579	0.128000	-0.000803	0.128803
580	0.000000	0.179368	0.179368
581	0.000000	0.201097	0.201097
582	0.343250	0.537474	0.194224
583	0.000000	-0.000803	0.000803
584	0.000000	-0.000803	0.000803
585	0.000000	-0.000803	0.000803
586	0.000000	-0.000803	0.000803
587	0.000000	-0.000803	0.000803
588	0.000000	-0.000803	0.000803
589	0.000000	-0.000803	0.000803
590	0.000000	-0.000803	0.000803
591	0.336000	0.318156	0.017844
592	0.000000	0.189791	0.189791
593	0.000000	-0.000803	0.000803
594	0.000000	-0.000803	0.000803
595	0.000000	-0.000803	0.000803
596	0.482333	0.207663	0.274670
597	0.000000	-0.000803	0.000803
598	0.000000	-0.000803	0.000803
599	0.125000	0.011867	0.113133
600	0.000000	-0.000803	0.000803
601	0.000000	-0.000803	0.000803
602	0.000000	-0.000803	0.000803

603	0.392333	0.502660	0.110327
604	0.266000	-0.000803	0.266803
605	0.715000	-0.000803	0.715803
606	0.535000	0.513833	0.021167
607	0.358000	0.460825	0.102825
608	0.000000	-0.000803	0.000803
609	0.000000	-0.000803	0.000803
610	0.176000	0.259441	0.083441
611	0.172000	-0.000803	0.172803
612	0.217000	0.187863	0.029137
613	0.000000	-0.000803	0.000803
614	0.152000	0.373735	0.221735
615	0.613000	0.508340	0.104660
616	0.000000	-0.000803	0.000803
617	0.000000	-0.000803	0.000803
618	0.000000	-0.000803	0.000803
619	0.000000	-0.000803	0.000803
620	0.000000	-0.000803	0.000803
621	0.624333	0.397808	0.226525
622	0.105000	0.117058	0.012058
623	0.000000	-0.000803	0.000803
624	0.000000	-0.000803	0.000803
625	0.000000	-0.000803	0.000803
626	0.000000	0.175765	0.175765
627	0.000000	-0.000803	0.000803
628	0.000000	-0.000803	0.000803
629	0.167000	-0.000803	0.167803
630	0.953000	0.525677	0.427323
631	0.209000	0.175765	0.033235
632	0.402500	0.419550	0.017050
633	0.000000	-0.000803	0.000803
634	0.702333	0.525885	0.176448
635	0.169250	0.062749	0.106501
636	0.000000	-0.000803	0.000803
637	0.000000	-0.000803	0.000803
638	0.268333	0.199168	0.069165

639	0.000000	-0.000803	0.000803
640	0.000000	-0.000803	0.000803
641	0.843667	0.519663	0.324004
642	0.000000	-0.000803	0.000803
643	0.000000	-0.000803	0.000803
644	0.131000	0.142612	0.011612
645	0.000000	-0.000803	0.000803
646	0.000000	-0.000803	0.000803
647	0.770000	0.653852	0.116148
648	0.000000	-0.000803	0.000803
649	0.463000	0.284667	0.178333
650	0.000000	-0.000803	0.000803
651	0.844500	0.408328	0.436172
652	0.000000	-0.000803	0.000803
653	0.218000	-0.000803	0.218803
654	0.000000	0.004714	0.004714
655	0.462000	0.356541	0.105459
656	0.000000	-0.000803	0.000803
657	0.416667	0.239033	0.177634
658	0.000000	-0.000803	0.000803
659	0.452667	0.277355	0.175312
660	0.000000	0.530139	0.530139
661	0.000000	-0.000803	0.000803
662	0.168667	0.175765	0.007098
663	0.000000	-0.000803	0.000803
664	0.000000	-0.000803	0.000803
665	0.000000	-0.000803	0.000803
666	0.000000	-0.000803	0.000803
667	0.000000	-0.000803	0.000803
668	0.000000	-0.000803	0.000803
669	0.236500	0.177913	0.058587
670	0.152000	-0.000803	0.152803
671	0.000000	-0.000803	0.000803
672	0.000000	-0.000803	0.000803
673	0.000000	-0.000803	0.000803
674	0.000000	-0.000803	0.000803

675	0.475333	0.212343	0.262990
676	0.144000	0.123870	0.020130
677	0.000000	-0.000803	0.000803
678	0.000000	-0.000803	0.000803
679	0.000000	-0.000803	0.000803
680	0.000000	-0.000803	0.000803
681	0.148250	0.141211	0.007039
682	0.000000	-0.000803	0.000803
683	0.334000	-0.000803	0.334803
684	0.000000	-0.000803	0.000803
685	0.122000	-0.000803	0.122803
686	0.000000	-0.000803	0.000803
687	0.163000	-0.000803	0.163803
688	0.000000	-0.000803	0.000803
689	0.190000	0.231368	0.041368
690	0.000000	-0.000803	0.000803
691	0.000000	-0.000803	0.000803
692	0.121667	-0.000803	0.122470
693	0.186250	0.162225	0.024025
694	0.120667	0.045379	0.075288
695	0.000000	-0.000803	0.000803
696	0.135000	-0.000803	0.135803
697	0.000000	-0.000803	0.000803
698	0.000000	-0.000803	0.000803
699	0.292000	0.302039	0.010039
700	0.000000	-0.000803	0.000803
701	0.000000	-0.000803	0.000803
702	0.000000	-0.000803	0.000803
703	0.401000	0.411998	0.010998
704	0.000000	-0.000803	0.000803
705	0.000000	-0.000803	0.000803
706	0.000000	-0.000803	0.000803
707	0.513250	0.039825	0.473425
708	0.284000	0.287442	0.003442
709	0.000000	-0.000803	0.000803
710	0.447333	0.547609	0.100276

711	0.188333	0.189364	0.001031
712	0.000000	0.187916	0.187916
713	0.171250	0.188044	0.016794
714	0.000000	-0.000803	0.000803
715	0.195000	0.191603	0.003397
716	0.158500	0.066789	0.091711
717	0.000000	-0.000803	0.000803
718	0.000000	-0.000803	0.000803
719	0.353000	0.387420	0.034420
720	0.000000	-0.000803	0.000803
721	0.000000	-0.000803	0.000803
722	0.000000	-0.000803	0.000803
723	0.156000	-0.000803	0.156803
724	0.203000	0.187583	0.015417
725	0.000000	-0.000803	0.000803
726	0.211250	-0.000803	0.212053
727	0.000000	-0.000803	0.000803
728	0.000000	0.226803	0.226803
729	0.000000	-0.000803	0.000803
730	0.000000	-0.000803	0.000803
731	0.166000	-0.000803	0.166803
732	0.000000	-0.000803	0.000803
733	0.000000	-0.000803	0.000803
734	0.000000	-0.000803	0.000803
735	0.584333	0.584201	0.000132
736	0.484000	0.238904	0.245096
737	0.000000	-0.000803	0.000803
738	0.000000	-0.000803	0.000803
739	0.363250	0.491260	0.128010
740	0.000000	0.279644	0.279644
741	0.000000	-0.000803	0.000803
742	0.287500	-0.000803	0.288303
743	0.000000	-0.000803	0.000803
744	0.000000	-0.000803	0.000803
745	0.000000	-0.000803	0.000803
746	0.000000	-0.000803	0.000803

747	0.000000	-0.000803	0.000803
748	0.000000	-0.000803	0.000803
749	0.455000	-0.000803	0.455803
750	0.000000	-0.000803	0.000803
751	0.000000	-0.000803	0.000803
752	0.000000	-0.000803	0.000803
753	0.130000	-0.000803	0.130803
754	0.000000	-0.000803	0.000803
755	0.000000	-0.000803	0.000803
756	0.103000	-0.000803	0.103803
757	0.300250	0.185769	0.114481
758	0.388333	0.059464	0.328869
759	0.000000	-0.000803	0.000803
760	0.000000	-0.000803	0.000803
761	0.000000	-0.000803	0.000803
762	0.115333	-0.000803	0.116136
763	0.000000	-0.000803	0.000803
764	0.000000	0.175765	0.175765
765	0.000000	-0.000803	0.000803
766	0.000000	-0.000803	0.000803
767	0.234000	0.284994	0.050994
768	0.124667	-0.000803	0.125470
769	0.000000	-0.000803	0.000803

TABLE B.1 Tableau présentant les résultats du réseau de neurones sur le jeu de test.

Annexe C

Photos de la mise en place de la station de la digue de l'Artha



FIGURE C.1 Photo d'un agent du CD64 en train de réaliser la saignée sur la digue à l'aide d'un marteau piqueur.



FIGURE C.2 Photo de deux agents du CD64 en train de remettre en état l'endroit où sera placé la station.



FIGURE C.3 Photo de deux agents du CD64 finalisant la mise en place des caissons sur lesquels les plaques des capteurs vont être fixées.



FIGURE C.4 Photo des deux plaques sur lesquelles les capteurs sont fixés.



FIGURE C.5 Photo de la mise en place de la gaine et tirage des câbles des capteurs.



FIGURE C.6 Coulage du béton sur la gaine électrique où passent les câbles des capteurs.



FIGURE C.7 Mise en place de la plaque horizontale de capteurs, vue de haut.



FIGURE C.8 Mise en place de la plaque verticale de capteurs, vue de haut.



FIGURE C.9 Mise en place de la plaque verticale de capteurs.

Annexe D

Fiche technique du capteur de pression



TRANSMETTEURS DE PRESSION PIEZORESISTIFS

PRESSIONS ABSOLUE, RELATIVE ET DIFFERENTIELLE

Ces transmetteurs piézorésistifs sont destinés aux applications en environnement industriel, en milieux liquides et gazeux, dans la gamme de 0,2 à 1000 bar.

Série 23 : Raccord fileté mâle

Série 25 : Membrane affleurante

Une Gamme Complete

Modèles absolus, absolus scellés, barométriques, relatifs, relatifs pour dépressions, différentiels liquide-liquide. 13 étendues de mesure nominales. Sortie courant ou tension.

Flexibilité

La conception modulaire de cette gamme autorise une grande flexibilité de réalisation et permet d'offrir à l'utilisateur un produit adapté à ses besoins. Plusieurs centaines de variantes sont réalisables: interfaces mécanique et électrique, matériaux, huile de remplissage, service oxygène, performances spéciales.

Qualité

Ces transmetteurs utilisent le capteur Keller série 10, de haute sensibilité, dont plusieurs millions d'exemplaires sont en service dans le monde. Chaque capteur est sévèrement contrôlé en pression, température et stabilité.

La désignation des transmetteurs est définie par le mode de fonctionnement et la position du zéro électrique correspondant à une pression nulle :

PAA-23/25 Pression absolue, zéro au vide

PA-23/25 Pression absolue, zéro scellé à 1000 mbar absolus

PR-23/25 Pression relative

PD-23 Pression différentielle

SERIE 23 / 25



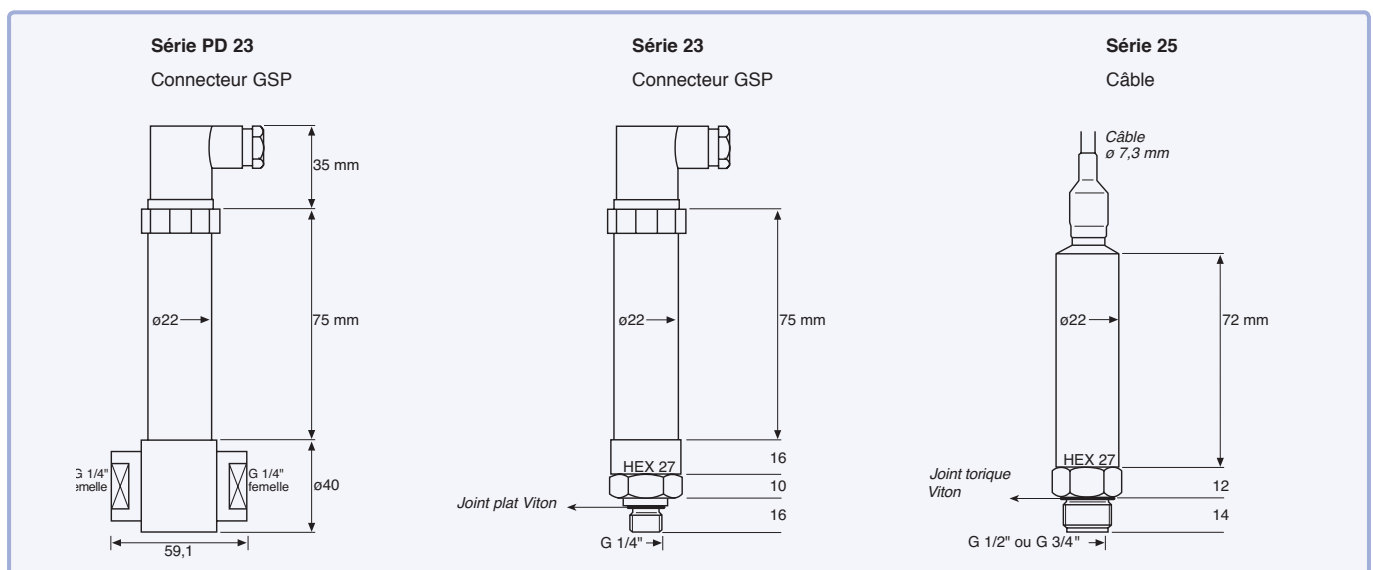
Série PD-23



Série 23



Série 25





SPÉCIFICATIONS

	Etendues de Mesure (EM). Surpressions. En bar										
PR-23	-1	-0,5	-0,2	0,2	0,5	1	2	5	10	20	
PR-25	-1	-0,5		0,5	1	2	5	10	20		
PD-23				0,2	0,5	1	2	5	10	20	
PAA-23				0,2	0,5	1	2	5	10	20	
PAA-25					0,5	1	2	5	10	20	
PA-23/25						1	2	5	10	20	50 100 200 400 600 1000
Surpressions	-1	-1	-1	2,5	2,5	3	4	10	20	40	100 200 300 600 900 1100
PD, surpressions côté -				1	1	2	3	5	7	10	
PD, pression statique	200 bar max.										

Précision*	±0,2 %EM typ.	±0,5 %EM maxi	*Linéarité (meilleure droite basée à zéro) + hystérésis + reproduction
Température de stockage/service	-40...100 °C		
Température compensée	-10...80 °C		
Effet de la température...	...sur le zéro, en %EM/°C		...sur la sensibilité, %/°C
· EM > 1 bar	0,005 typ.	0,01 max.	0,05 typ.
· EM 1 bar	0,010 typ.	0,02 max.	0,05 typ.
· EM 0,5 bar	0,015 typ.	0,03 max.	0,05 typ.
· EM 0,2 bar	0,025 typ.	0,05 max.	0,05 typ.
Stabilité typique à long terme			
· EM > 2 bar	0,1 %EM typ.	0,2 %EM max.	
· EM ≤ 2 bar	1 mbar typ.	2 mbar max.	

Signal de sortie	4...20 mA / 2 Fils	0...10 V / 3 Fils
Alimentation (U)	8...32 Vcc	13...32 Vcc
Charge de ligne (R _Q)	(U - 8 V) / 0,02 A	> 5 k

Raccordement électrique
- Connecteur GSP EN 175 301-803-A (DIN 43650)
- Par câble avec évent (PR).

Respiration < 0,1 mm³
Raccord pression Voir dessins au recto
En contact avec le fluide mesuré Acier inoxydable type AISI 316L, Viton®
Protection, CEI529 IP65
Masse PR/PA/PAA: ≈ 180 g. PD: ≈ 440 g
Isolation > 100 MΩ@50 V
Durée de vie > 10 millions de cycles 0...100 %EM à 25 °C
Fréquence limite 1 kHz (5 kHz en option)

Options

Raccord pression. Sortie électrique Consulter, nombreuses possibilités
Etendues de Mesure Autres EM comprises entre 0,2 et 1000 bar
Précision (Linéarité) 0,2 %EM max. Spécifications utilisateur sur demande
Matériaux en contact avec le fluide Nombreuses possibilités
Huile de remplissage Huile fluorée (service O₂), huile d'olive, huile basse temp.
Protection IP67, IP68 (version câble)
Pression statique plus haute jusqu'à 600 bar (seulement PD)

RACCORDEMENT ELECTRIQUE

Transm. 2 fils	Transm. 3 fils	Connecteur GSP	Connect. Binder 723	Câble
-	GND	1	1	blanc
OUT/GND	OUT	2	2	rouge
+Vcc	+Vcc	3	3	noir

