Faculty of Bioscience Engineering

Feature extraction and classification of sonar echoes from rigid and elastic targets

Application to underwater mines detection and environmental risk mitigation in marine ecosystems

Author	Struyf Nathan
Supervisors	Lambot Sébastien (UCLouvain)
	Hurtado E. Camilo A. (UCLouvain/Royal Military Academy)
Readers	Lopera Tellez Lucia Olga (Royal Military Academy)
	Craeye Christophe (UCLouvain)
Academic year	2025 - 2026

Thesis presented in fulfillment of the requirements for the degree of Bioengineer in environmental science and technology, specialized master's programme

Index Terms - Classification and pattern recognition, machine learning classification of acoustic signals, surveillance and threat assessment techniques, active sonar, elastic and rigid scattering, feature extraction.

Acknowledgments

The first person that I would like to thank is Camilo Andress Hurtado, for the constant help that he provided to me throughout my work on the thesis. Camilo has always been available to answer my questions and organized meetings on a regular basis to check on my progress, with the help of useful tools facilitating follow-up. I also thank Pr. Lambot for the confidence that was put into my work and for accepting to supervise my master thesis. Finally, I want to thank all the people who helped me externally throughout my work, either by supporting me through tough times or by helping me with the writing and improvements of my work.

Contributions

Table 1: Author contributions and use of artificial intelligence tools.

Contribution	Nathan Struyf	Sebastien Lambot	Camilo A. Hurtado	AI	AI uses
Conceptualisation	X	X	X		
Methodology	X		X		
Investigation	X			X	Coding assistance in Python and MATLAB.
Formal analysis	X				
Validation	X	X	X		
Writing – original draft	X		X		
Writing – review and editing	X	X	X	X	Grammar and spelling corrections.
Visualisation	X			X	Improvement of table formatting and.
Supervision		X	X		

AI was used only as technical assistance (coding, grammar correction and table formatting). It was not used to generate scientific content, interpretations, results or conclusions. The generative AI tools used are ChatGPT and Claude. No AI was ever used for redaction purposes. All written texts are mine and proofreading and error corrections were performed by relatives and supervisors.

Contents

List of Figures

List of Tables

List of Acronyms

SAS Synthetic Aperture Sonar

SSS Side Scan Sonar

SAR Synthetic Aperture Radar

IFFT Inverse Fast Fourier Transform

FFT Fast Fourier Transform

FRFT fractional Fourier Transform

DFT Discrete Fourier Transform

HMM Hidden Markov model

SNR Signal-to-Noise Ratio

ADC Analog-to-Digital Converter

AUV Autonomous Underwater Vehicle

ROV Remotely Operated Vehicle

UXO Unexploded ordnance

AUC Area Under Curve

ROC Receiver Operating Characteristic

HDPE High Density Polyethylene

FOV Field Of View

MDF Medium Density fiber

DCNN Deep Convolutional Neural Network

DC Direct Current

LFM Linear Frequency Modulation

UNCLOS United Nations Convention on the Law Of the Sea

CCW Convention on Certain Conventional Weapons

SLAMF Système de Lutte Anti-mines Marines Futur

USLB Ultra-Short Base Line

GPS Global Positioning System

ToF Time of Flight

CFAR Constant False rate Alarm

TPR True Positive Rate

FPR False Positive Rate

TNR True Negative Rate

CI Confidence Interval

Chapter 1

Introduction

1.1 General context

Underwater mines and UXO (UneXploded Ordnance) are one of the numerous environmental threats in oceans and seas around the world. The difference between a mine and a UXO is that a UXO is a broader term that refers to any military munition that was fired or dumped at sea and did not detonate [1], while an underwater mine is a more specific term, referring to a weapon used to destroy vessels or submarines. They are among the last remnants of past wars that are still a threat to this day. In the North Sea, they are mostly residual mines from World War I, with the North Sea Mine Barrage, for instance [2] being one of the conflicts that led to some important accumulation of UXO's, where over 70,000 naval mines were dropped in the sea, between the Orkney Islands (an archipelago in Scotland) and Norway. World War II was also a time were lots of naval mines were used. The countries that were involved developed new types of bombs, such as magnetic naval mines, invented by the Germans [3]. Not only in the North Sea are those objects a threat. There are also mines and bombs in other parts of the world, from numerous conflicts across the globe, such as the Korean War [4] or the Russo-Japanese War [5], but also military tests [6]. As a matter of fact, underwater explosives have been used long before World War I. They were already invented by the Chinese in the 14th century [7]. At that time, those bombs were made of wood and sealed with putty, some kind of malleable sealing material that was frequently used at the time. This design of bombs gained popularity in Europe and quickly reached its coasts by the 17th century [8]. Their presence in the water has been a problem for centuries. Underwater mines are mostly invisible threats that are often very difficult to reach as they are located in some of the harshest marine conditions that can be encountered around the globe. The North Sea is actually well known for being one of the most hostile seas on earth because of its violent winds all year long, its shallowness, making it one of the most turbulent seas in the world. Moreover, storms are frequent and visibility can be low throughout the year because of frequent fog, making navigation difficult [9]. However, it is necessary to locate and disarm those bombs as they are a direct threat to many marine species, human life, human settlements and off-shore constructions [10] and the environment in general. Even in today's conflicts underwater mines are still used. The most recent example for this is the conflict in Hormuz strait, where the Iranian government is believed to have placed underwater mines as stated by US Secretary of State Marco Rubio [11]. Some come close to harbors and fishing spots, they are therefore in close contact with human settlements and activities everyday [6]. For example, in the years following the great wars, it was not so uncommon for dredging vessels to be entirely destroyed after capturing explosives in their fishing nets. For instance, a Ukrainian dredging boat was completely destroyed and sunk on a Danube canal in 2025, as reported by the Maritime executive [12].

Underwater bombs can be found in many shapes, materials and in many different types of environments. These characteristics should be carefully taken into account for the management of underwater mines. This is thus a difficult task, because the models and algorithms used by specialists to detect and recognize the mines have to take all these points into account.

These mines can be of spherical, torpedo like and so on and reach sizes of more than a meter in diameter and several hundred kilograms [13]. These explosives, that can still be actively charged, could be harmful to marine species. In fact, their explosive charge may remain active for decades after their release [14]. Another important aspect to take into account refers to the level of burial depth of these ordnances. Some can be in free field, attached to the seabed by chains (called moored mines). Some of them are drifting, but this is quite uncommon. The others are found on the seabed. Either proud (which means that they are on the surface with no level of burial), partially, or fully buried. In these conditions, they are the most difficult to spot and locate with sonar technologies because they are hindered.

1.2 Objectives

1.2.1 General objective

The objective of this thesis is to perform classification with machine learning algorithms on a dataset of acoustic data from sonar analysis of targets, based on their physical properties. More precisely, on their tendency to behave as either elastic or rigid reflectors with respect to the acoustic waves sent by the sonar. The effectiveness of a classifier can be assessed by evaluating its ability to place the targets in the right category. In the global pipeline of UXO's and naval mines disposal and neutralization, classification is one of the most critical parts. The whole efficiency of this process relies on the ability of experts to discern dangerous ammunition from harmless objects. The key aspects are thus to categorize all harmless objects that were detected as harmless and, more importantly, not missing any bomb by placing it in the wrong category. In either case, the consequences would be negative. If a harmless object is classified, the minesweepers would reach it for nothing. But if a bomb is categorized as harmless, it would remain in the environment and continue to pose a threat. To perform classification with any machine learning algorithm, data is required to train and test them. In this study, the data comes from a Study conducted at the University of New Hampshire, in the United States [15]. In this study, they collected acoustic data for several target types, insonified by a sonar in air, for several types of environments of varying complexity. The data would then undergo a series of processing and extraction steps to build a new dataset, suitable for classification. Four target types are separated into two categories, elastic and rigid, according to their material type.

1.2.2 Secondary objectives

- Understand the global concepts and technologies of sonar as a whole. As it is a completely unknown field, there are many concepts that need to be studied before using the data properly.
- Understand the dataset and assess in what way it is useful in the field of classification. How can it be used for UXO and mine remediation.
- Use other types of classification algorithms to compare their effectiveness and to validate the results of the main classification algorithm. The first classifications are performed with supervised machine learning, but the second type of classifier that is used is logistic regression. The purpose of this is mainly to serve as a comparison point with supervised machine learning algorithms.

- Assess the cause of potentially poorer classification performance and evaluate the origin of the problem. Finding a solution, even though it would require processes that would not be applicable in real conditions, could nevertheless yield valuable information and guide future researchers on the topic in better understanding the problem.

1.3 Cautions with the data

It is important to note that for this study the data used were recorded in air. The usual application would be conducted in water, but this data set was chosen because it combines several attributes that make it stand out from other datasets. First, several types of targets were under study. two spheres that behave as elastic targets and two cardboard letters that behave more like rigid targets. Second, the experimental conditions were heavily controlled. The targets were placed in several types of environments, mimicking real objects found in the seas. Key environmental parameters were measured in order to guaranty rigorous and precise results in the calculation of the speed of sound for example. Finally, the experiments were reproduced several times for every possible combination of target type and environment type. Blank measurements were also recorded for all environments. This allows a better understanding of the acoustic behavior of the experimental setup. Thanks to all these points, the data set is large, with precise and ready to use acoustic data that is well suited for classification purposes.

Chapter 2

State of the art

An efficient way to detect UXO's and mines is by using sonar technologies. It consists of sending sound waves toward a specific target or background and collecting the reflected sound wave. The reflection carries information about the target that can be exploited either on the raw signal itself for signal analysis or by reconstructing an image with different techniques of beam focusing. Sonars are analogous to radars but differ in the type of wave used. Radars use electromagnetic waves, more precisely radio frequencies, whereas sonars emit beams of sound waves toward a target and record the returning echo. The sound wave requires a medium to travel, whereas an electro-magnetic wave does not. Using a beam of light for the detection of underwater targets is not feasible because light waves tend to attenuate rapidly in water. For above water uses, radars are the best option, because air is a low-density medium, with sparsely distributed particles in a gaseous state. The waves can propagate easily, while being slowed down and dissipated only by a marginal amount. In water, electromagnetic waves are much more affected by dissipation by the fluid through two processes: conversion of energy into heat, and scattering by the molecules. The energy of the wave decreases as it progresses through the fluid and is therefore not usable under these conditions [16]. To use light-based devices, one would need to use very low frequencies which are more penetrating, but the antenna would have to be too large for practical use, because the smaller the frequency, the larger the wavelength and thus the larger the antenna [17].

Sonars come in a variety of technologies. The simplest one is the passive sonar. It essentially works as a microphone that listens to its surroundings, in the same way a camera works. Active sonars differ in the way that they emit a sound wave with what is called an emitter (basically a speaker). And the echo is retrieved by a receiver (same hardware as for a passive sonar) The emitter and the receiver of a sonar are together called transducer. A sonar that emits a broad fan-shaped beam and forms multiple receive beams is called a multibeam sonar [18], which is used to build bathymetric maps (maps of the seafloor). If both the transmitter and the receiver of the sonar are located in the same place, the sonar is called monostatic. But these two devices can be physically separated. In this case, the sonar is called bistatic. Furthermore, a sonar can have multiple receivers, in order to get more angular information on the insonified objects for one burst of sound. Sonars are also used for 3D imagery. The choice of frequency is important according to the purpose of the study. For 3D imagery, precision is required and higher frequencies are used for higher image resolution. But high frequencies do not penetrate solid media at all, in contrast to lower frequencies that penetrate from centimeters up to more than a hundred meters for the lowest frequencies (sub bottom profiler sonar), in the seabed, through sand or even rocks for instance [19]. Therefore, lower frequencies are more suited for target location and classification than for imaging, because they would provide very low resolution images in contrast to high frequencies.

One common method used for acoustic imaging is the Side Scan Sonar (SSS). The sonar sends a conical or fan-shaped beam that has an angular tilt giving a grazing angled view of the insonified scene.

SSS works by stitching together slices of echoes recorded along the track of the ship transporting the sonar. The sonar is typically mounted in a device called a "towfish" that is towed behind the vessel [20]. Another way to make images with acoustic data is to use SAS processing. SAS is analogous to Synthetic Aperture Radar (SAR) (the technology used to recreate high resolution satellite images), but with sound. In SAS imaging, which is a technology used since the late sixties [21], the sonar sends several beams along a trajectory, at a constant rate. The image is constructed by attributing a grid of pixels to the scene and the contribution of every beam of sound reaching a given pixel is summed , which is a step called focusing, allowing to improve the resolution to build high resolution images [22].

While collecting data is necessary, another important step is to classify the collected data, in order to separate and categorize the objects present in the scanned environments. One must identify the type of ordnance before choosing a disposal method, and it is also essential to distinguish bombs from rocks or between unexploded and exploded bombs for instance. Classification can be done on several parameters regarding the bombs. If the goal is to distinguish bombs from rocks, then the physical properties will be used in the classification algorithms. But if the type of the bomb is the studied parameter then the shape of the targets will be the discriminant factor for classification.

There are several ways to perform classification. But mainly two types of algorithms are used. On one hand there is machine learning and on the other hand there is deep learning. Both are powerful tools that help researchers achieve their objectives. They share the same working principle. A model learns from data by iteratively making predictions and adjusting its parameters to minimize errors. Machine learning algorithms can be divided in two main groups. The first group is the supervised classification algorithms. Supervised algorithms are those for which the answer to the classification problem is known beforehand. For instance, a classifier that needs to classify images of cars on the color of these cars. For each image of a car, a label indicating the car's color is provided, and the model is trained using this labeled information, enabling the model to predict the outcomes for unseen data. Unsupervised machine learning trains on datasets lacking labels. They essentially work by recognizing underlying patterns and similarities in the data to make predictions. A common application of such models is to group people according to their behaviors, such as making customer groups based on their purchasing habits [23] [24]. Deep learning algorithms use artificial neural networks in several layers in between the input and output layers. They are fast and effective algorithms but require a lot of input data to train the model, which is still a common limitation as lots of datasets are too small to efficiently train deep learning algorithms [25]. Before the rise of machine learning and deep learning, other classification methods already existed. Stochastic classification [26], Rule based and fuzzy rule based classification [27] are some examples.

2.0.1 Rigid vs elastic targets

The composition of the bombs is also an important factor to keep in mind as the response to the sound beam differs accordingly to it. Objects targeted by the sonars can act as rigid or elastic reflectors, and this has to do with the composition of the materials used to make the ordnances and their internal structure. Elastic objects vibrate when they are excited by sound waves By sending a burst of sound with decreasing (or increasing) frequency, which is called a Linear Frequency Modulation (LFM), often called a chirp based on the sound emitted by birds - the target is more prone to vibrate at its fundamental or multiple harmonics, which means that it will emit amplified scattering waves that will appear in the echo and these signatures can be used for classification purposes. For the purpose of illustration, it is analogous to someone rubbing a humid finger on a crystal glass filled with water, when the resonance frequency is reached, the glass begins to resonate and emit sound. A rock, for instance, shows very limited vibrations in comparison to more elastic objects, whereas a metal sphere vibrates more under the influence of an acoustic wave and is thus considered as an elastic target.

2.0.2 ka coefficient calculation

A key parameter in the characterization of echoes, or scattering waves is the k.a coefficient. The formula to calculate it is given in equation 3.3 and takes into account the wave length and the characteristic length of the object (the radius for a sphere). Basically, this number compares the size of the object to the size of the wavelength and dictates whether resonance is possible. It tells the diffusion regime of a sound wave or, and it tells whether an object acts as a resonant reflector or not. Depending on the value of this coefficient, three regimes exist. Rayleigh regime (k.a « 1), elastic or Mie regime (k.a = 1) or geometric regime (k.a » 1) [28]. If the object is smaller than the sound wave (Rayleigh regime), the target will be seen as a unique point and resonance will be very low. With comparable sizes (elastic regime), resonances are possible and the internal and material characteristics of the targets are present in the scattering waves of the echoes. When the size of the object is far greater than the wave length (geometric regime), specular reflections are dominant. This is comparable to the reflection of light on an object. Resonant waves are less perceptible because their energy is smaller than the specular reflection or they can even be absent. With a chirped signal, there is a window of k.a (from the lowest frequency to the highest). Which is interesting because the smaller value is closer to an elastic regime, while the highest value of k.a approaches the geometric regime. So, depending on the regime in the studied environment, it is possible to know whether resonance waves are possible and therefore perform classification on such characteristics with the given dataset. In the end, the ka coefficient tells whether an object will produce internal reflections, and thus whether the data are suitable for classification between elastic and rigid targets [29].

2.1 Pipeline for UXO management

As explained previously, data classification is a key process in the management of UXO in waters around the world. In addition, countries around the world are subject to different guidelines and laws.

For instance, coastal states are subject to the general obligation to protect their marine environments. The basis of this obligation is the convention of the United Nations Convention on the Law Of the Sea (UNCLOS), signed in Jamaica in 1982. This convention states that these countries have to manage coastal pollution. Although marine bombs are not explicitly stated in this convention, these countries have, therefore, a margin of discretion with respect to the disposal of the bombs [30].

Protocol V (on Explosive Remnants of War) of the 1980 Convention on Certain Conventional Weapons (CCW), states that states have an obligation to dispose of explosive remnants from past conflicts, in addition to cooperation between different states in the disposal activities and strategies. But this protocol has its limitations as it is targeted to more recent conflicts, while UXOs in the North Sea are mostly remnants of the two World Wars [31].

The European Union also acts towards this direction through its "Marine Strategy Framework Directive". The objective of this framework directive is to reach a good ecological state of marine waters, where UXOs are considered as a risk for marine ecosystems and a chemical poisoning source. Concerned countries must carry out several tasks, involving the mapping of at-risk zones, monitoring the impacts, and setting up countermeasures such as UXOs neutralization [32].

Managing UXOs therefore requires following a series of steps.

2.1.1 Under water survey

This whole pipeline of steps starts with the research of the mines in the water, which involves any kind of detection device. In this case, a sonar is the obvious choice as explained in part 1.1 of this document. The simplest setup involves a hull-mounted sonar. The vessel surveys a designated perimeter, with the sonar emitting impulses that will bounce off of the seabed, the targeted mines, and any other natural or artificial object found in the water. The echoes are then collected by the receiver of the sonar. European countries have made several efforts regarding underwater mines and UXOs management technologies. For instance, A French initiative led by the "Ministère des armées et des anciens combattants" Système de Lutte Anti-mines Marines Futur (SLAMF) [33], an acronym standing for future marine mines countermeasures system in English. This initiative aims to modernize french infrastructures needed for marine mines management, by replacing ancient survey boats with robotized systems for instance. A fleet of Autonomous Underwater Vehicle (AUV)s and Remotely Operated Vehicle (ROV)s was recently deployed to detect and neutralize bombs in the North Sea. These vehicles contain sonar technologies allowing to detect the bombs in environments that are difficult to access for manned missions and allow for tactical and quick surveys of vast surfaces of waters.

2.1.2 Location

The next step is the location of the targets. This step is ensured with a Global Positioning System (GPS), which is connected to the sonar. But there is one problem. Telecommunications do not work underwater for reasons explained earlier, therefore a GPS can not receive any positioning signal if it is under a layer of water. The GPS must be attached to the boat and then connected to the sonar. This is a problem if a AUV is used for detection, because it is not physically connected to any surface object yet still requires location. Several solutions exist, such as inertial navigation. The AUV bears motion and rotation sensors to compute its position in real time. Its main advantage is that it is completely autonomous but, as a drawback, its precision decreases over time [34]. Another method is the Ultra-Short Base Line (USLB) [35]. It works as follows: a boat sends a ping to the AUV, that will then answer to this ping by sending a ping to the boat. Then the distance between the boat and the AUV is calculated with the time of flight of the acoustic wave and the angle of the wave. This distance can then be converted into a position using the GPS coordinates of the vessel. When the issue of positioning is resolved, every time a ping of sound waves is emitted, the position data is stored jointly to the impulse, So when a mine is detected and correctly classified as such, its precise location is easy to retrieve. Location of underwater objects is a completely separate field with its own challenges, that is necessary in the pipeline of UXO.

2.1.3 Classification

The next step is to find the mines within the sonar data. After the data is collected, it has to be fed to a classification algorithm. The data can be processed beforehand to gain information about the targets. For example, by conducting SAS processing. Classification is performed on one or several properties of the targets. These properties have to be chosen adequately for the classification to be as efficient as possible. This step is fundamental as it is here that one can tell the difference between a bomb and something else that isn't harmful. Those properties are intrinsic to the detected objects, such as the shape, the materials, the hollowness, etc.

2.1.4 UXO management

If the targets are successfully located and classified as UXO's or any kind of naval mine, they have to be identified and then neutralized, by using the right deactivation methods by professionals. There are two possible ways to neutralize a bomb: in-situ and ex-situ management. In-situ is when the bomb is being dealt with and detonated in the water. A first method is to neutralize the bomb by deactivating its detonation device. Another method is to proceed to a controlled detonation of the bomb, by placing an explosive and detonating it remotely, after making sure that nothing can get hurt within the explosion range [36]. The use of ROV helps a lot for this purpose, as such devices can be precisely guided to several mine sites, can be mounted with cameras for visual navigation and can even be mounted with explosives to remotely stick to bombs. A concrete example of such ROV is the "Poisson Autopropulsé", One of France's initiatives to remotely dispose of dangerous bombs [37]. One last example of method is to displace the entire bomb out of the water and managing it elsewhere, where it is safer to deal with, although this method is quite dangerous, it can be used for ordnance with decayed explosive charges that are not a threat anymore. When dealing with a bomb, it is important to know its nature and structure to be able to choose the most appropriate disposal method.

2.2 Classification between rigid and elastic targets

The idea of classification of targets based on parameters such as elasticity/rigidity is motivated by the fact that UXO's and mines are mostly made of metallic shells. These materials behave as elastic targets, meaning that they can deform and recover their original shape when subjected to a constraint, such as an incoming sound wave produced by an active sonar [38] [39]. These vibrations, which are at the microscopic scale, generate waves that scatter to the environment. These waves can be retrieved by the sonar's receiving end, and the echo will contain acoustic signatures from the targets. Rigid targets, such as rocks, generally vibrate way less than metallic objects such as bombs and mines. The incoming sound wave from a sonar reflects on the surface of the target, causing only specular reflection to appear in the echo, while the echo of an elastic target contains a mix of both specular reflections and internal vibrations scattering waves. Therefore, the study of the echoes of the different targets makes it possible to differentiate them on the basis of their physical properties, without requiring visual differentiation. Natural underwater objects such as rocks are easily confused with dangerous objects with visual approaches. So, the advantage of classification on physical properties through the use of sonars is to get rid of the confusion caused by the potential visual resemblance between rocks and UXO's by studying and comparing their intrinsic properties.

However, most current studies use data from high-frequency sonars (>100 kHz) to reconstruct highly detailed images of the studied scenes. But the problem with this approach is that high frequency sound waves penetrate neither the seabed nor the targeted objects at all, making it difficult to analyze them when they are partially or fully buried in the sand. Low frequencies (around 10 kHz) are more suitable for this purpose, as they have better penetrating abilities. Even better, a chirped signal, with modulated frequency, allows to gather more information on the targets as explained in section 1.1.2. In fact, frequencies lower than 25 kHz take advantage of target structural acoustic features, robustness to target motion effects, long-range propagation, and enhanced sediment penetration [40].

Several studies have used the targets' physical properties to support classification or identification. For instance, Dmitrieva et al. [41] identified the shell and filler materials of spherical shells from the time delays of secondary reflections in wideband sonar echoes, without requiring supervised training data. Paper: Material recognition based on the time delay of secondary reflections using wideband sonar pulses. Other works have exploited the resonance-related features of elastic shells for parameter estimation, including estimation of the radii from resonance peak intervals in both free-field and proud-on-sediment configurations [42]. More recently, neural-network-based approaches have also been applied to sonar target classification. Xu et al. [43] proposed a one-dimensional convolutional neural network trained on simulated backscattering morphological functions to classify the material of elastic spherical shells and estimate their radius and thickness. Lee et al. used cepstral features extracted from real active-sonar data, including power-normalized cepstral coefficients, as inputs to a convolutional neural network for target-versus-clutter classification [44]. Although these approaches achieved promising results, their reliance on deep neural networks may make the physical interpretation of the discriminative features less direct for future studies than in methods explicitly based on scattering mechanisms.

Zerr et al. [45] used elastic properties combined with aspect dependence to improve the classification of mines. The resonance scattering analysis they conducted consisted in looking at different types of periodic phenomena in the scattering wave captured by the sonar, significantly contributing to the target's echo. They explained that these characteristics are visible in both the time and frequency domains. In the time domain, it is visible as several small echoes following the specular echo, while for the frequency domain, it corresponds to frequency domain modes. They concluded that the analysis of resonance scattering has excellent potential for classification purposes.

Jia et al. [46] stated that target echoes contained rigid and elastic components that could be separated and that the rigid component contained information on the shape and the size of the targets, while the elastic component contained information on the elastic nature of an object, which in turn depends on the material and internal structure of the object. To separate the components, they used the fractional Fourier Transform (FRFT) and some filtering steps. They concluded that the separation of elastic acoustic components had consistent theoretical features that can lay the foundation for classification. They emphasize several key features present in the elastic component which describe the composition of the targets' composition. They also concluded that some of the features, extracted at a specific range of frequencies, were able to discern objects of the same shape but from different compositions, especially in time domain analysis. These results show the potential of key feature extraction for target classification based on their physical properties.

Dasgupta et al. [47] proposed a class-based target identification framework using multi-aspect acoustic time-series data. Rather than training a classifier for each individual target, they grouped targets into classes whose members shared similar scattering physics, arguing that this strategy is more suitable for real applications where training data are not available for every possible target. Features were extracted from sequences of transient backscattered waveforms and used to train a Hidden Markov model (HMM) classifier, designed to exploit the aspect-dependent evolution of the target response. The method was evaluated using both computed and measured acoustic scattering data from submerged elastic targets. In the synthetic case, classification was performed among four classes of elastic ellipsoidal shells differing mainly in length and Youngs modulus. In the experimental case, the problem was formulated as a binary classification between a class of five elastic shell targets with similar external shape but different internal structures, and a set of non-target or clutter objects, including metallic, plastic, rock, drum, and wooden objects. The study considered both noise-free and noisy conditions, but all data were free-field; the authors concluded that class-based classification is promising, while extension to more realistic propagation environments remained necessary.

Seo et al. [48] studied underwater cylindrical-object detection using active sonar time-series data rather than sonar imagery. Their approach was formulated as a binary classification problem between target-present and no-target/clutter cases, with the experimental evaluation focusing on an aluminum cylinder as the target and rock measurements as clutter. Features were extracted from the matched-filtered received signals by applying theDiscrete Fourier Transform (DFT) and using the magnitude spectrum as the input feature vector. A logistic regression classifier was then trained to discriminate target responses from clutter, providing a simple and interpretable machine-learning baseline. The study combined simulated and experimental data: simulated signals were generated using a finite-cylinder scattering model and a seabed clutter model, while real data came from PondEx10 measurements. In the simulation study, the classifier achieved an AUC of about 0.90, while in the experimental transfer case, where simulated data were selected and used for training, the model reached an average AUC of 0.93, with a detection rate of approximately 0.77 and a false-alarm rate of approximately 0.06. The results suggest that spectral features from matched-filtered sonar echoes can support targetclutter discrimination and that simulation-based training can be useful when experimental data are limited, although further work is needed to improve robustness under more realistic environmental variability.

Although previous studies have shown that acoustic scattering information can be useful for underwater target detection and classification, relatively few works have explicitly addressed the problem as a rigid-versus-elastic target classification task. Existing approaches often focus on different but related problems, such as multi-aspect shape classification from sonar imagery, separation of rigid and elastic scattering components, class-based identification of elastic targets, or target-versus-clutter detection using spectral features. These studies demonstrate the relevance of geometric, elastic, and spectral information, but they do not directly evaluate whether time-series features extracted from backscattered echoes can discriminate rigid from elastic targets across environments of increasing complexity. This distinction is particularly relevant for mine-countermeasure applications, where man-made objects such as UXOs or mine-like targets may exhibit elastic scattering mechanisms, while natural clutter such as rocks is often dominated by rigid-like scattering. In addition, several previous studies are restricted to free-field or simplified conditions, leaving open the question of how classification performance changes when the target response is affected by interfaces, roughness, or partial burial.

2.3 Governing equations

2.3.1 Elasticity

An elastic material is able to regain its original shape after being exposed to a constraint. In other words, the deformations of such objects are called elastic deformations. Many materials behave as elastic objects, such as metals [49] or even polymers [50]. The reason behind this elasticity varies. At the atomic level, metals look like a three-dimensional lattice with metallic bounds between every atom. When a stress is applied, the whole lattice changes in size and shape. When the constrain disappears, the lattice regains its previous size and shape, at the lowest energy state [49]. For plastic materials like rubber, elasticity is due to the extension of the polymer chains [51].

Mathematically, such deformations are explained as follows: For linear elasticity, which concerns small deformations, the elongation is proportional with respect to the force applied to the object. The elastic modulus, or Young modulus (E) as seen in equation 2.1, is a constant number which is a property of materials and it quantifies the relation between stress (σ) and strain or deformation (ε). The units of the Young modulus are Pascals, which makes it homogeneous to a pressure tensor. More precisely, it is the mechanical stress that would cause a one hundred percent elongation, or, in other words, doubling the length of the object under stress. Rigid materials have, for instance, a high Young modulus, going up to several gigapascals. Elastic materials have a lower Young modulus, as low as several kilopascals (Mathematical theory of elastic structures [52]).

$$E = \frac{\sigma}{\varepsilon} \tag{2.1}$$

$$\varepsilon = \frac{l - l_0}{l_0} \tag{2.2}$$

$$\sigma = \frac{F}{S} \tag{2.3}$$

With ℓ_0 = original length of the object, l = length after elongation due to stress, F = force of applied strain and S = surface on which the strain is applied.

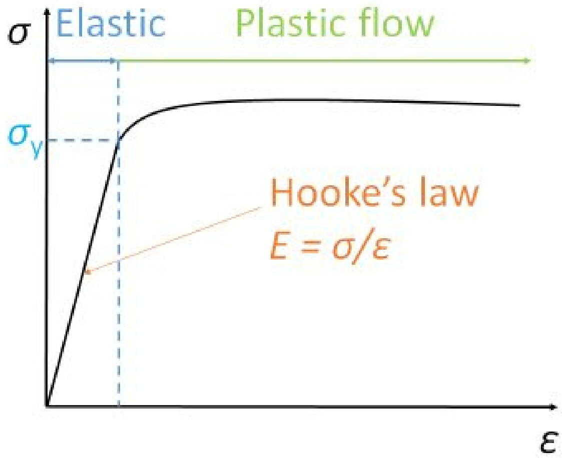

Figure 2.1: Graph of Hooke's law. The linear part of the graph represents the elastic behavior of a material. Graph taken from the DoITPoMS website, from university of Cambridge. [53]

This figure 2.1 and equation 2.1 represent Hooke's law. In the graph, the elastic behavior is in the beginning of the curve and, for limited values of strain, corresponding to the area in the function where the slope is constant. After that value of (ε), the behavior tends to be plastic like (non linear relationship between stress and strain) and where deformations are permanent. According to this graphic, deformations remain elastic for a limited strain, perfect for application to stresses caused by a sound wave that are of micrometric scale. it is important to note that Hooke's law is a simplified representation valid for uniaxial deformations.

2.3.2 Acoustic waves

An acoustic wave is the propagation of a variation of pressure. In sonar applications, the pressure is generated by the vibration movements of the emitter membrane. In other words, it is a propagation of a perturbation, with alternating compressions and expansions. Individual water, air or any fluid molecules near the source of the sound wave oscillate locally, transmitting their energy to neighboring molecules, and so on, until the wave hits a target or dissipates over time. Molecules do not travel along the sound wave. Equation 2.4 describes the propagation of the pressure wave in a fluid, with p being the pressure, c the speed of sound in the fluid, and t the time. ∇^2p is the laplacian of the pressure. It takes the second time derivative of the pressure in the three space dimensions [54].

$$\nabla^2 p - \frac{1}{c^2}\frac{\partial^2 p}{\partial t^2} = 0 \tag{2.4}$$

For an acoustic wave in a fluid, only longitudinal waves (waves parallel to the energy transfer direction) can propagate because the fluids are not subjected to shear forces (a shear force acts parallel to the surface of the material). In solids, shear forces appear. So, when an acoustic wave hits an object, a portion of the transmitted energy is reflected off of the surface, and the rest is absorbed by the object, causing internal vibrations and new types of propagation waves to appear. First, P waves (which stands for Primary waves), which are in the same direction as the incoming acoustic wave. The second wave that appears is the shear force wave or S wave (which stands for secondary wave), because solid objects are subject to shear forces, in contrast to fluids. S waves appear only when the incoming sound wave is not perpendicular to the surface of the object. It is the angular tilt that induces shear in the solids [55]. The energy of these two waves are transmitted back to the fluid, but only in the form of P waves. An acoustic wave can be physically described with the wave equation, which includes the pressure and density of the fluid, as well as its propagation speed and time. What it means for the study of targets is that elastic objects will have a lot of internal vibration in the form of P and S waves that will be transmitted to the water and limited reflection. for rigid objects, it tends to be the opposite.

The global energy of an acoustic wave hitting a target can be summarized as follows:

$$incident = reflection + transmission + energy\ losses \tag{2.5}$$

$$transmission = S_{waves} + P_{waves} \tag{2.6}$$

It is the proportion of reflection and transmission that essentially explains the rigid or elastic behavior of the targeted objects.

2.3.3 Reflection and transmission coefficients

As explained in the previous section, when an acoustic wave hits an object, there are two possibilities. The incoming wave is reflected back to the environment, and it is absorbed by the object. Both can occur at the same time too. In the first case, it is called reflection, and in the second case, the waves that are absorbed reflect in the material as P and S waves and the energy is transmitted back to the surrounding fluid, which is called transmission. Both reflection and transmission are the result of intrinsic properties of every material and can be quantified with a corresponding coefficient [56]. Another possible variation of the path of the sound waves is refraction, which occurs when the path of the beam is slowed or accelerated by a variation in the medium, such as temperature or density. This is much less concerning in air than in water, especially when the path of the wave is short. But if the sound wave propagates in water, refraction is a phenomenon to be taken into account because sea waters are stratified because of three phenomena: static pressure, salinity and temperature [55]. Refraction causes bending of the wave's trajectory [55].

Equations 2.7 and 2.8 represent the reflection and transmission coefficients of an acoustic wave:

$$R = \frac{E_{reflected}}{E_{incident}} \tag{2.7}$$

$$T = \frac{E_{transmitted}}{E_{incident}} \tag{2.8}$$

R = reflection coefficient and T = transmission coefficient and E is the energy of the wave, incident and transmitted.

Both coefficients are linked and depend on the energy of the incoming wave. In the case of a planar acoustic wave, the coefficients can be written as functions of the acoustic impedance (Z) [56], with the relation in 2.9:

$$Z = \rho c \tag{2.9}$$

With c being the speed of sound in the fluid or the object and ρ the density of the fluid or the object.

The coefficients then become [56]:

$$R = \frac{Z_2 - Z_1}{Z_2 + Z_1} \tag{2.10}$$

$$T = \frac{2Z_2}{Z_2 + Z_1} \tag{2.11}$$

Z_1 is the impedance of the fluid and Z_2 the impedance of the object.

And, to take into account the fact that P and S waves are generated, new coefficients appear. R_P for the reflected P waves, T_P for the P waves in the solid and T_S for the S waves in the solid, when incident P waves from the fluid that are absorbed by the solid become P and S waves within the solid.

The amount of reflection is conditioned by the difference in impedance between the medium and the object: the impedance determines the scattering properties of a wave [57] and tell how much a material "resists" sound waves, analogously to the electric resistance of materials towards a current. Air, for instance, has a low impedance, while solid objects have a high impedance. If the values of both impedances differ by a large amount, there will be more reflection. Water has a higher impedance that is closer to the impedance of solid objects which means that there will naturally be more internal reflections than in experiments conducted in air. This means that the medium in which the study is conducted has important impacts on the classification [58]. Equations 2.10 and 2.11 show this behavior. If both impedances of the medium and the target are similar, reflections will be close to 0 and most of the energy will be transmitted. But if the impedances differ a lot, more reflections will appear.

2.3.4 Linear Frequency Modulation - chirp signal

An important characteristic of signals used in sonar applications is the modulation of incident signals. A sonar can be used by sending a simple, single-frequency wave. But by modulating the frequency, for instance by generating an increase or a decrease of the pulsed signal's frequency over time, it is possible to get more details from the targeted objects because they are insonified by a variety of frequencies, exciting them in different ways. It is also useful for its important pulse compression potential for better resolution with long range reaches [59]. Such signals are called chirps, in reference to the sounds produced by birds and are used by birds in nature as well as cetaceans like dolphins, which roam their environments and hunt by sending complex chirped signals. A chirped signal that has a rising frequency is called an up-chirp, and when the frequency decreases, it is called a down-chirp. Figure 2.2 shows a linear up-chirp, with a constant amplitude.

Equation 3.1 represents a linear chirp [60]:

$$u(t) = \mathrm{rect}\left(\frac{t}{T}\right) \exp\left(j\pi \frac{B}{T} t^2\right) \tag{2.12}$$

where T is the pulse duration, B is the swept bandwidth, and the rectangular window is defined as

$$\mathrm{rect}\left(\frac{t}{T}\right) = \begin{cases} 1, & |t| \le T/2 \\ 0, & \text{otherwise} \end{cases} \tag{2.13}$$

The instantaneous frequency of the chirp follows directly from the phase derivative:

$$f_i(t) = \frac{1}{2\pi} \frac{d}{dt}\left(\pi \frac{B}{T} t^2\right) = \frac{B}{T} t \tag{2.14}$$

Here, $u(t)$ denotes the complex envelope of the chirp signal, t is time, T is the pulse duration, and B is the swept bandwidth of the signal. The ratio B/T corresponds to the chirp rate (or sweep rate), i.e. the speed at which the instantaneous frequency increases over the duration of the pulse. The quadratic term t^2 inside the exponential is what produces this linear variation of frequency with time once the phase is differentiated, as shown in equation 2.14. The $\mathrm{rect}(t/T)$ factor simply restricts the signal to the interval $[-T/2, T/2]$, giving it a finite duration. By sending a broad range of frequencies with the sonar, the insonified objects are excited in such a manner that they can show more resonance patterns, which is useful in echo analysis. Furthermore, a chirped signal provides a better temporal resolution after pulse compression in the signal processing steps, and the signal to noise ratio is also improved.

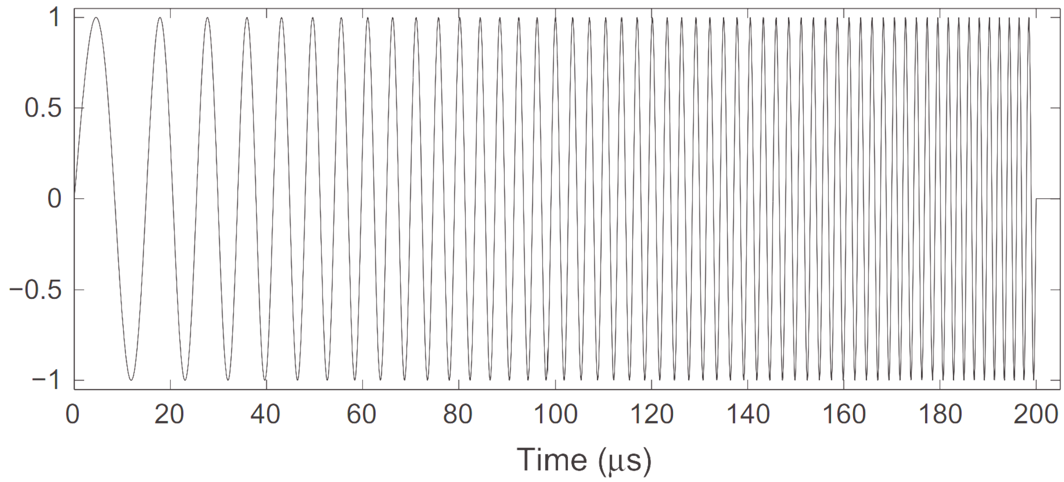

Figure 2.2: Linear up-chirp signal with a constant amplitude. The graph was taken from "Chirp excitation of ultrasonic guided waves", (Michaels et Al.). The x axis represents the time and the amplitude of the signal is on the Y axis (arbitrary units to be adapted for the type of signal). [61]

2.4 Differences between elastic and rigid targets

In this section, the differences between elastic and rigid targets will be presented. When an echo reaches the receiving end of a sonar, it causes the membrane of the receiver to vibrate. this vibration is then transformed into an analog signal so it can be interpreted by a computer. The value of the raw signal is thus expressed in volts. Rigid targets tend to only reflect the incoming sound wave when insonified by a sonar and it shows in their echo. But for targets behaving more like elastic objects, internal vibrations caused by the structure of the object itself to bend and squeeze under the pressure of the sound wave are visible in the scattered wave spectrum through small echoes appearing after the main specular echo. The next two figures (2.3 and 2.4) are those of signals that have been processed and several filters were applied. Units on the y axis are, after processing, arbitrary and do not correspond to any existing physical quantity (this reads volts because the raw data collected by the transducer is a tension, but the tension no longer makes sense after the data processing (The chain of signal processing will be discussed in a future section of this document). The x-axis corresponds to the Time of Flight (ToF) of the signal. These are windows extracted around one specific ToF corresponding to the location of the target within the signal.

2.4.1 Echo of elastic targets

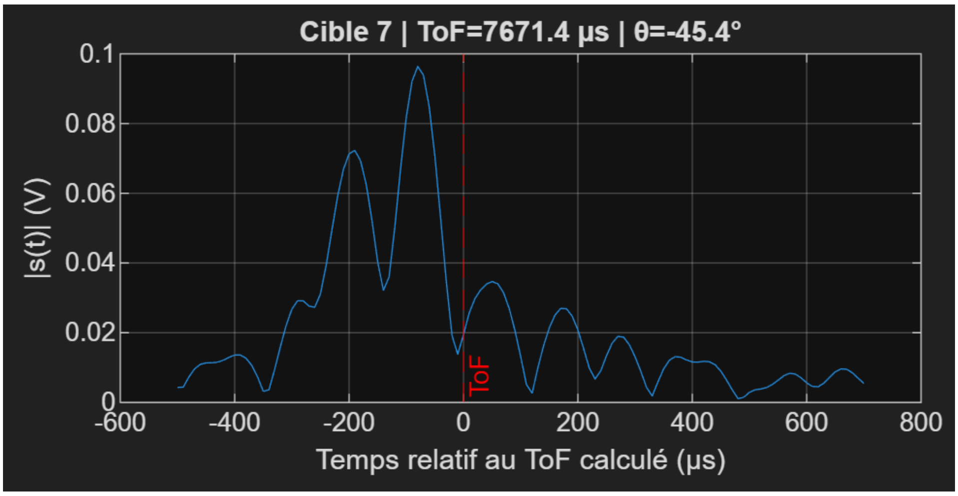

Figure 2.3: Typical echo of an elastic target. Graph created with MATLAB with the code provided by Blanford et al. (2024)
[15]

As explained earlier, signals from elastic targets show smaller echoes after the main specular reflection, resulting from the wave reflected from the surface. Those smaller echoes are visible on the figure after 0 on the x axis. It is also visible that the intensity of these echoes tends to decrease. At around 450 to 500 microseconds, noise from the background begins to appear. Therefore, the echoes tend to last for around 450 to 500 microseconds before vanishing.

2.4.2 Echo of rigid targets

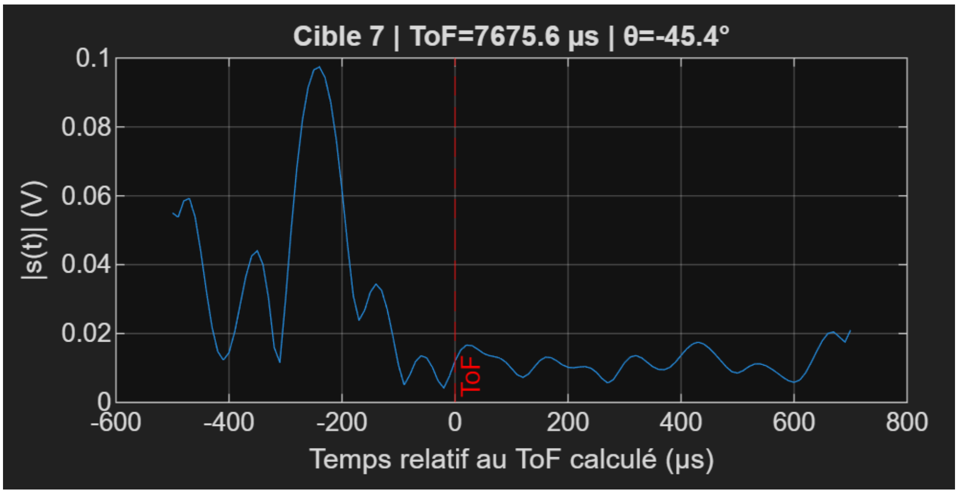

Figure 2.4: Typical echo of a rigid target. Graph created with MATLAB with the code provided by Blanford et al. (2024)
[15]

The typical echo of a rigid target displays only a specular reflection, as expected. The smaller waves positioned at the tail of the function that do not exceed 0.02 units are the contribution of the noise from the background. It is important to note that both echoes were extracted from different targets, but the environments were the same for both at the same place in the scene, same experimental environment (position number 7, - see figure 3.3 - on flat interface. The elastic target is the aluminum sphere and the rigid target is the letter O).

2.5 Processing chain

The reflection and transmission waves are collected by the receiver of the sonar and converted to an analog signal. This signal is still raw and needs to be processed. The reason behind this is that the raw data are not directly exploitable for several reasons. The first reason is that the data are passing through some hardware, wires, and several electronic devices. This distance, even though small, causes a non negligible delay to the whole signal. Therefore, the signal has to be shifted to recover its correct position. Then, the data can be passed through some filters, which can help extract more information on the signal, to make some parts of the signal easier to analyze, and to filter out some of the signal that is outside a given frequency band. Furthermore, the signal can be subject to transformations like the Hilbert transformation, which generates the analytic representation of the signal (the complex valued signal). A matched filter can also be applied to the signal to improve the signal to noise ratio. All these processing steps will be developed in more detail in a further chapter.

Processed data can be used for several purposes, such as SAS processing, whose operation was already explained in chapter 1. Although SAS processing might be an interesting approach to perform classification, there is a fundamental problem to classify using the physical properties of the targets. Those properties are perceptible in the time series, or echoes of the targets, but the synthetic aperture process involves the focusing of the data into single values for the pixels of the scene, because the objective of SAS is to reconstruct a high-resolution image with acoustic data. But this focusing step completely removes the time series, and thus all the information relative to the echoes due to internal vibrations useful for classification between rigid versus elastic targets is lost. The fact that the time series are completely lost is related to the type of SAS processing technique used, as there are multiple ways to perform SAS processing.

Chapter 3

Material and method

3.1 Description of the in-air SAS dataset

3.1.1 Data acquisition

To perform classification on different types of targets, data is required. The data used in this study were collected by a research team from the University of New Hampshire (USA), where they analyzed four types of targets, in four different controlled environments and in air [15]. The four targets, as shown in 3.1, are two spheres that have the same diameter of 10.2 cm. The first sphere is whole and made from polyurethane, and the other one is hollow and made from aluminum. the other two targets are two letters, an O and a Q made from MDF (medium density fiber), a type of material made of compressed wooden fibers. Their thickness is 1.91 cm, and their diameter is 20.3 cm.

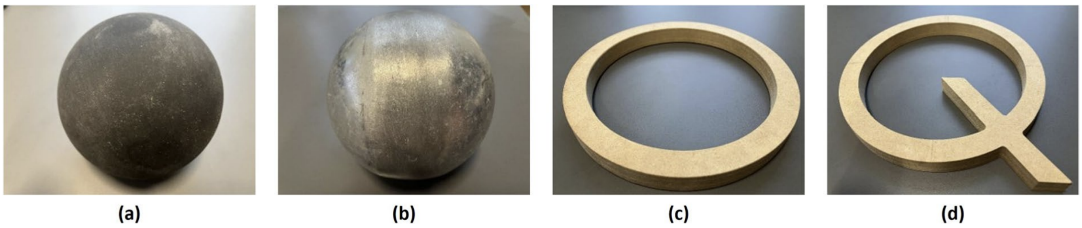

Figure 3.1: The different types of targets used for the experiments. (a) is the whole polyurethane sphere, (b) is the hollow aluminum sphere, (c) is the MDF O and (d) is the MDF Q. Images taken from Blanford et al. (2024)
[15]

The first two targets exhibit strong internal resonances and can be placed in the elastic category, while the latter two show very limited internal vibrations, with specular reflection being the predominant characteristic. They are thus placed in the rigid category.

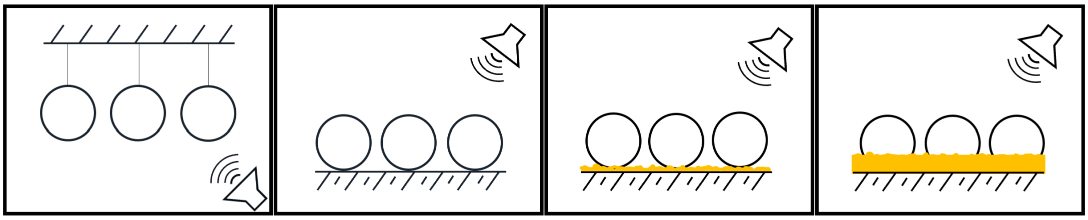

Figure 3.2: The different environments under study in the experiments. (a) is free-field, (b) is proud on a flat interface, (c) is proud on a rough interface and (d) is partially buried in a rough interface, in air.
Self-made figures.

Figure 3.2 shows the different environments under study. The complexity of the environments increases [15], with the last two being what is closest to real life conditions, adding reflections, diffuse back-scattering (which is noise generated by the diffusion of the sound waves by the particles of the surface), and shadowing (that occurs when an object in the scene blocks the sound wave, creating a signal void, or black area, behind it). The rough interface is made of plastic High Density Polyethylene (HDPE) pellets [15]. This type of plastic is used extensively in the food industry for packaging and bottles, for example because of its rigidity and resistance to high and low temperatures and the fact that it is easily recycled.

All environments were tested for every target type. Data collection for a specific target can be summarized as such: for the free-field experiments, seven targets of the same type for every measurement were hung with fishing wires, always at the same places and heights and the sonar was placed below the targets, facing them upward at an angle of $+25°$ [15]. For the other three environments, seven targets, still of the same kind, are placed within virtual boxes on a table measuring five meters by two meters [15], as shown in figure 3.3, and those boxes are always at the same location for every repetition of the experiments. A sonar is placed on the very edge of the table, facing the targets downward, with an angle of $-25°$ [15]. for all environments, the sonar sent a first chirped signal and then collected the echo. Then it translated five millimeters to the right to send a second ping of acoustic signal [15]. In total, for one scene, 1,001 pings were acquired (one ping every five millimeters over a five-meter track) [15]. The reason behind so many takes is to recreate a high resolution image of the whole scene. But for the purpose of classification based on the time series, so many pings is also an advantage because it allows to have a lot of signals from different angles of the same targets, allowing for the creation of a extensive dataset. Furthermore, for one scene of one given type of target and one given type of environment, several scans of 1001 pings were taken, with the targets being repositioned randomly within the virtual boxes, to get variations in the signals [15]. Figure 3.4 displays the number of collections for all variations of the experiments. for instance, the experiments where the targets were hollow aluminum spheres, placed on a flat interface, were repeated 34 times times, yielding 34 times 1,001 pings for this condition alone.'

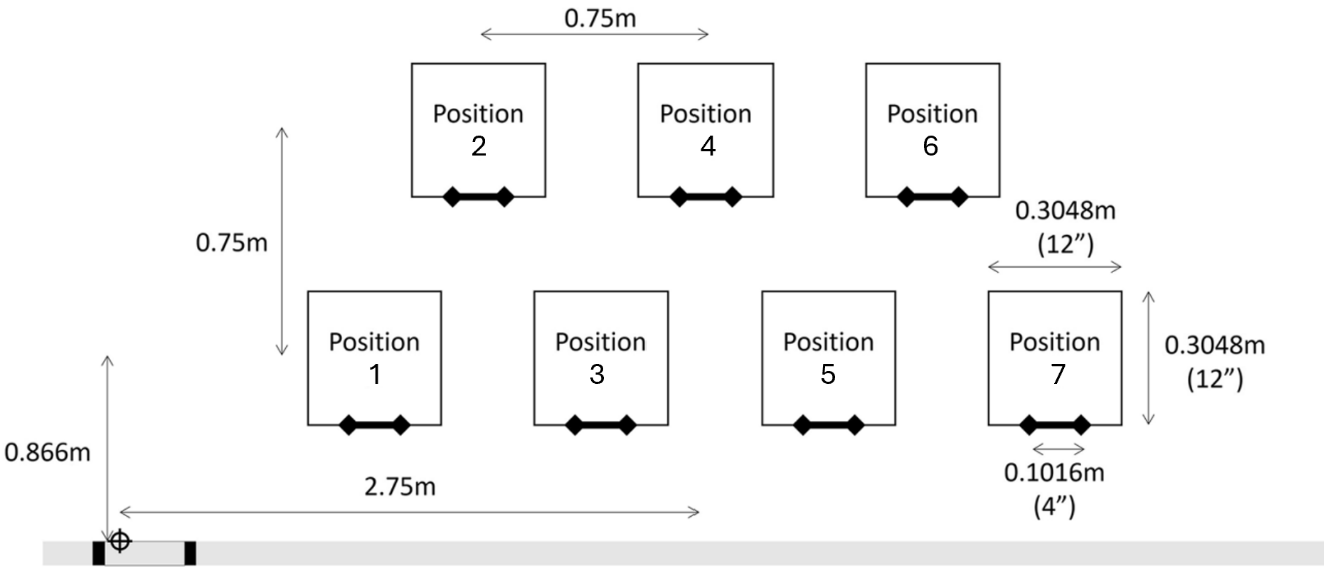

Figure 3.3: Layout of the scenes for flat and rough interfaces, with the targets always placed within the 30 cm by 30 cm virtual boxes. Figure taken from Blanford et al. (2024) [15]

Filename	Description	Number of Collections
t0e1_xx.h5	No target present, free-field	8
t1e1_xx.h5	Solid sphere, free-field	42
t2e1_xx.h5	Hollow sphere, free-field	32
t3e1_xx.h5	Letter O, free-field	33
t4e1_xx.h5	Letter Q, free-field	31
t0e2_xx.h5	No target present, flat interface	14
t1e2_xx.h5	Solid sphere, flat interface	34
t2e2_xx.h5	Hollow sphere, flat interface	34
t3e2_xx.h5	Letter O, flat interface	63
t4e2_xx.h5	Letter Q, flat interface	53
t0e3_xx.h5	No target present, rough interface	32
t2e3_xx.h5	Hollow sphere, proud on rough interface	32
t3e3_xx.h5	Letter O, proud on rough interface	32
t4e3_xx.h5	Letter Q, proud on rough interface	31
t1e4_xx.h5	Solid sphere, partially buried in rough interface	32
t2e4_xx.h5	Hollow sphere, partially buried in rough interface	32
t3e4_xx.h5	Letter O, partially buried in rough interface	36
t4e4_xx.h5	Letter Q, partially buried in rough interface	35
noise_xx.h5	Noise recording (no waveform transmission)	5

Figure 3.4: Number of collections for every experimental variations. Table taken from Blanford et al. (2024) [15]

The sonar used in the research has a limited Field Of View (FOV) of 120 [15]. and the down-chirp ranges from 30 kHz to 10 kHz [15]. The chirp used in this case is a little bit different from the general equation given in section 2.3.4.

$$u(t) = w(t)\sin\left(2\pi\left(f_1 + \frac{f_2 - f_1}{2t_p}t\right)t\right) \tag{3.1}$$

$$w(t) = \begin{cases} \frac{1}{2}\left[1 - \cos\left(\frac{2\pi t}{rT}\right)\right] & 0 \leq t < \frac{rT}{2} \\ 1 & \frac{rT}{2} \leq t \leq \left(1 - \frac{r}{2}\right)T \\ \frac{1}{2}\left[1 - \cos\left(\frac{2\pi t}{rT}\right)\right] & \left(1 - \frac{r}{2}\right)T < t \leq T \end{cases} \tag{3.2}$$

With r being the Tukey window fraction, which is equal to 0.1 in this context. The Tukey window is a special type of function called a window function, that is applied to a signal. The role of a window function is to smoothly taper the signal's beginning and end to zero or a low value, eliminating sharp edges and more importantly, fixing both ends of the signal at the same value, which is needed to apply a Fourier transform. This is done to avoid spectral leakage, a phenomenon where the energy of a frequency overflows on the neighboring frequencies in the frequency domain. This could lead to unwanted artifacts and calculation errors [62]. A window function contains a main lobe (containing most of the signal's energy) around the origin an two edge lobes. There are several types of window functions: the simpler ones are square or triangle functions. More complex windows include Hanning and Hamming windows. The Hanning window edges both ends to zero, while the Hamming puts them near zero and can be seen as an optimized Hanning window. The Tukey window is a special one. It consists in a combination of two windows. The center is a rectangular window and the edges are cosine lobes. this is variable. The advantage of the Tukey window, is that it has a flat center, which means that the tapering only affects the beginning and the end of the function that needs to be shaped. The Hanning and Hamming windows are essentially lobes that are continuously increasing, then decreasing, that shape the whole function. This is why the Tukey window was chosen, because it preserves the middle parts of the signal [63].

The notation in equation 3.1 lacks an imaginary part because it is a physical representation and not an analytical one as shown in section 2.3.4. The difference is that the chirp here contains a variable amplitude $w(t)$, which is the consequence of the Tukey window. Equation 3.2 is the Tukey window equation. It essentially causes the amplitude to be modulated. In this case, at first, the amplitude increases to a maximum value, for a time equal to $rT/2$ as seen in equation 3.2. Then, near the end of the signal, the amplitude decreases again for a time or a time equal to $(1 - r/2)T$. In short, the equation of $u(t)$ causes the frequency to decrease and the equation of $w(t)$ causes the amplitude to modulate.

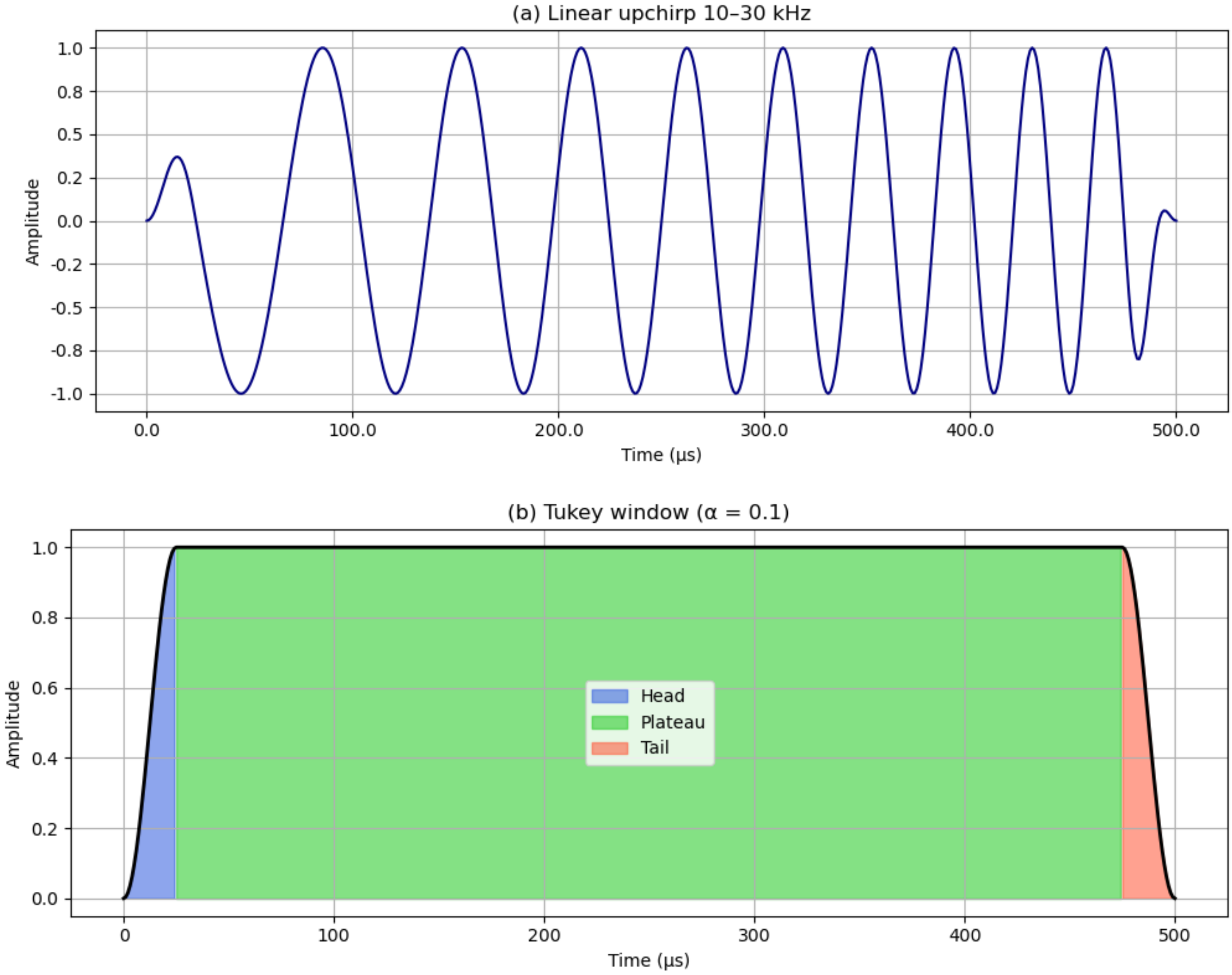

Figure 3.5: Linear up-chirp signal with a modulated amplitude (a), and the visualization of the Tukey window (b), used by Blanford et al. (2024) for the data. [15]

The modulated chirped signal is shown in figure 3.5 (a) and the beginning and end of the signal are tapered to 0, contrary to the classical chirped signal as shown in figure 2.2. The graph of the Tukey window is shown in figure 3.5 (a). The beginning and the end of the signal are tapered to zero, contrary to the classical chirped signal that is shown in figure 3.5 (b). The blue and red zones in 3.5 correspond to the zones where the amplitude is modulated and both last for 5% of the duration of the signal ($rT/2 = (0.1 * 500)/2 = 25\ \mu s = 5\%$ of $500\ \mu s$). The duration of this modulation is conditioned by the Tukey window fraction r and a higher value of r gives off higher modulation windows. The green zone is the plateau, where the signal remains unaffected.

In the article, the data was used to generate high definition acoustic images using SAS processing.

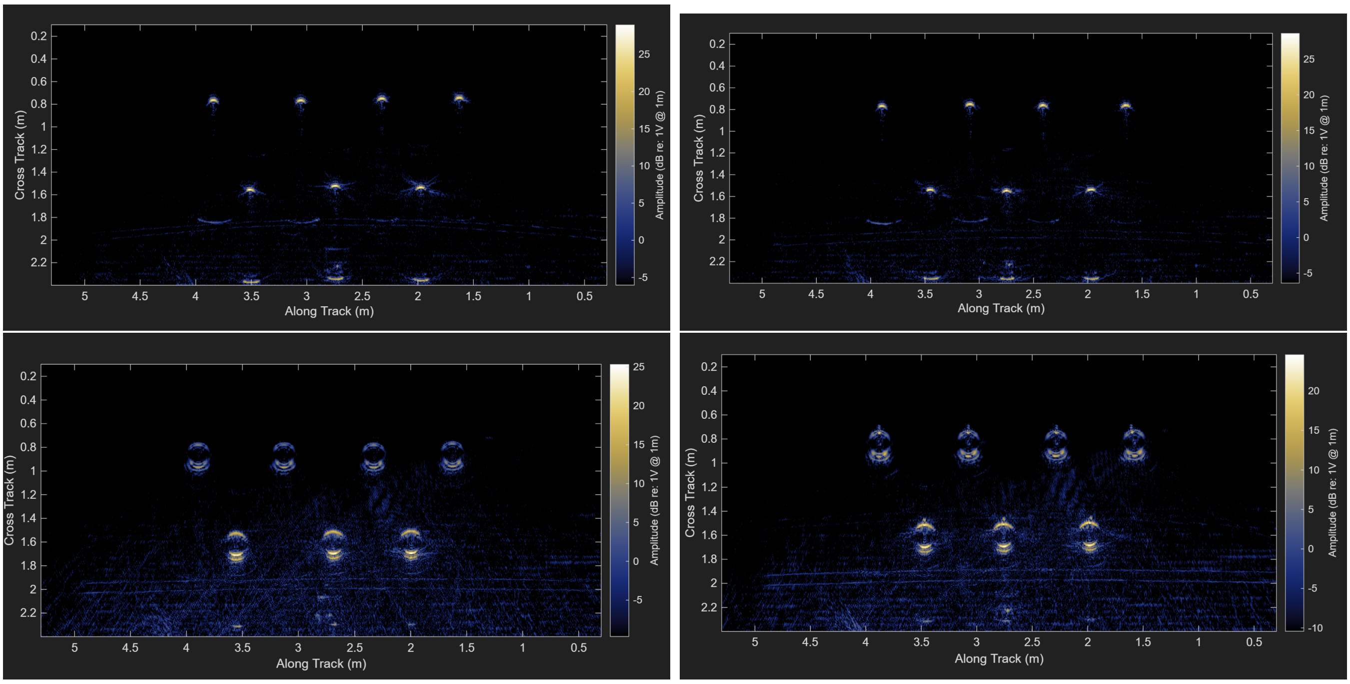

Figure 3.6: Reconstructed images from the acoustic data with synthetic aperture processing, in free field environment. Image reconstructed with the code provided by Blanford et al. (2024) [15]

In figure 3.6, upper left image is from the polyurethane sphere. The upper right image is from the hollow aluminum sphere. Both bottom images are from the MDF letters. On the left is letter O and on the right is letter Q. These images are quite good for visual analysis and classification, but lack information on the physical properties of the objects. It is therefore impossible with this method to differentiate two objects that have similar shapes but different compositions, for example. The shape of the spheres is not very well defined. But for both MDF letters their shape is clearly visible.Their overall shape is well defined, and the tail of the Q is also visible. Nevertheless, it is important to note that these are for ideal cases because there is close to no noise from the background. Images of the same targets in rough interface environments are much worse, as in figure 3.7.

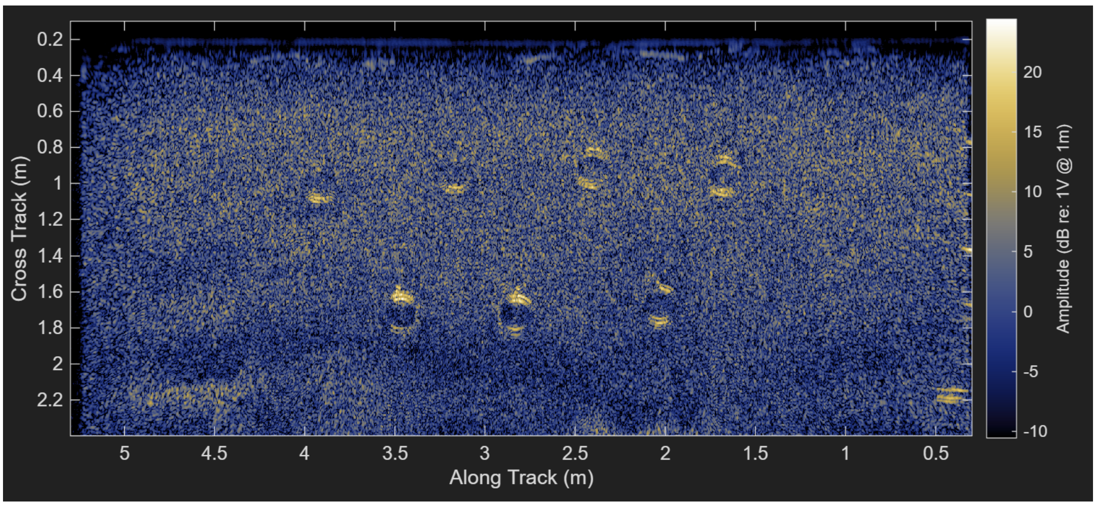

Figure 3.7: Reconstructed images from the acoustic data of letters Q, partially buried in rough interface. Image reconstructed with the code provided by Blanford et al. 2024 [15]

In figure 3.7, the noise level is high due to the reflections caused by the rough interface. In some places, the level of noise almost reaches the same magnitude as the reflections of the targets. Some of the targets are even barely visible, such as the two top left ones. Visual interpretation is here complicated because the characteristics are hidden behind noise. However, this situation is closer to real life conditions and many of the mines encountered in the seas would be partially covered or fully buried in sand or covered in vegetation. This is why other methods of target identification are needed if machine learning is used. Several studies that rely on SAS images use deep-learning techniques for classification. Williams et Al. [64] used convolutional neural networks to train on a limited set of SAS images. They insisted on the fact that large datasets are not always needed to perform classification using deep learning if the relation between training data and network capacity is optimized. This is achieved by employing small networks. Their results were promising, achieved quite good classification performances. Williams et Al. [65] also used Deep Convolutional Neural Network (DCNN) to classify on SAS sonar data, by taking mugshot images of scenes scanned by a sonar under water and after SAS processing. They compared several types of objects mimicking existing mines (manta-shaped, torpedo-like, cone-shaped) to rocks, each type of mine was classified with rocks, the classification was not performed with several types of mines together. Thus, these experiments use SAS images to classify targets using their shapes and relying on the ability of the classifiers to differentiate a rock from a target only based on the shape. But in a situation where a rock would have a shape very similar to a target, the classification could be wrong. This is why working with signals that encompass the target's physical properties along with information about their shape seems to be a good lead to perform classification.

When machine learning is used and not deep-learning with SAS image-based classification, other methods are coupled with them, showing that using SAS images only is not reliable. For example, Abu et Al. [66] used a combination of SAS images and optical imagery to discriminate man-made objects from natural objects. They transformed the optic images into SAS-like images, and they use a set of descriptors quantifying the relationship between the targets' shadow and highlight. Either by using lower frequencies that are more penetrating in sand or vegetation. but it would reduce the resolution, or by working directly on time series to extract signatures that are specific to target types and compositions.

Unlike for underwater mine detection, this data is collected by sending acoustic waves in air. At first glance, it might seem contradictory but it is not a problem, actually. because air and water are both fluids, the only thing that matters here is the presence of internal reflection waves. These appear when the system is in an elastic regime (as explained in chapter 1) if the calculated ka coefficients are close to 1, internal reflections might appear. This coefficient is calculated with equation 3.3 [67]. Equation 3.4 represents the link between the wavelength and the frequency.

$$ka = \frac{2\pi a}{\lambda} \tag{3.3}$$

$$\lambda = \frac{c}{f} \tag{3.4}$$

k ($= 2\pi/\lambda$), the wave number and a is the characteristic length (the radius for a sphere). For objects that have a more complex shape, such as the MDF letters, the characteristic length is not well defined and it changes with the aspect angle. λ is the wavelength. The signal being a chirp, there is a window of wavelengths between 10 kHz and 30 kHz [15]. For these two values, the ka coefficients are 2.14 and 6.4 respectively. At these values, internal resonance is possible and thus the data can be used for classification because the elasticity components are exploitable.

3.1.2 Getting the speed of sound in the air

The temperature was measured in two separate spots within the experimental setup, with the help of two k-type thermocouples, as well as the relative humidity owing to two probes. These variables were taken for every ping and the speed of sound was deduced from them using empirical relationships. An accurate air temperature measurement is critical because important parameters described later in this paper require an accurate value of the speed of sound [15].

Two cases are used to calculate the speed of sound. The first case is when the speed of sound is calculated only with the temperature, with this equation:

$$c(T) = 331.3 + 0.61\mathcal{T} \tag{3.5}$$

With $\mathcal{T}$ the average temperature, given by:

$$\mathcal{T} = \frac{\tau_1 + \tau_2}{2} \tag{3.6}$$

Equation 3.5 is derived from the speed of sound for an ideal gas given by [68]:

$$c(T) = \sqrt{\gamma R T} \tag{3.7}$$

With γ, the heat capacity ratio, which is 1.4 for the air [69]. R is the specific gas constant of the air, which is 287 $J.kg^{-1}.K^{-1}$ [69] and T the temperature. The development of this equation gives equation 3.5. This equation means that the higher the ambient temperature, the faster the speed of sound.

The second case uses the measured relative humidity (x_w) as well as the CO_2 fraction in the air (x_c) in a parametric equation from Cramer et al, 1992 [70].

$$\begin{aligned} c = a_0 + a_1 t + a_2 t^2 &+ (a_3 + a_4 t + a_3 t^2) x_w \\ &+ (a_6 + a_7 t^2 + a_8 t^2) p \\ &+ (a_9 + a_{10} t + a_{11} t^2) x_c \\ &+ a_{12} x_w^2 + a_{13} p^2 + a_{14} x_c^2 + a_{15} x_w p x_c \end{aligned} \tag{3.8}$$

Table 3.1: List of the values of the parameters from Cramer's equation.

Parameter	Value	Meaning	Parameter	Value	Meaning
p	1.01E+05	Pressure	a7	3.73E-08	Empirical
xc	0.0004	molar fraction of CO_2	a8	-2.93E-10	
a0	331.5024	Empirical	a9	-85.2093	
a1	0.6030		a10	-0.2285	
a2	-0.0005		a11	5.91E-05	
a3	51.4719		a12	-2.8351	
a4	0.1495		a13	-2.15E-13	
a5	-0.0007		a14	29.1797	
a6	-1.82E-07		a15	0.0004	

3.2 Classification

With the data explained, it is time to dive into the processing flow that is applied to the data. The first steps are used to realign the signal to its true place after it was delayed by passing through the hardware. The other steps involve filtering to enhance the signal, improve the signal to noise ratio, and remove unwanted noise.

3.2.1 Preprocessing

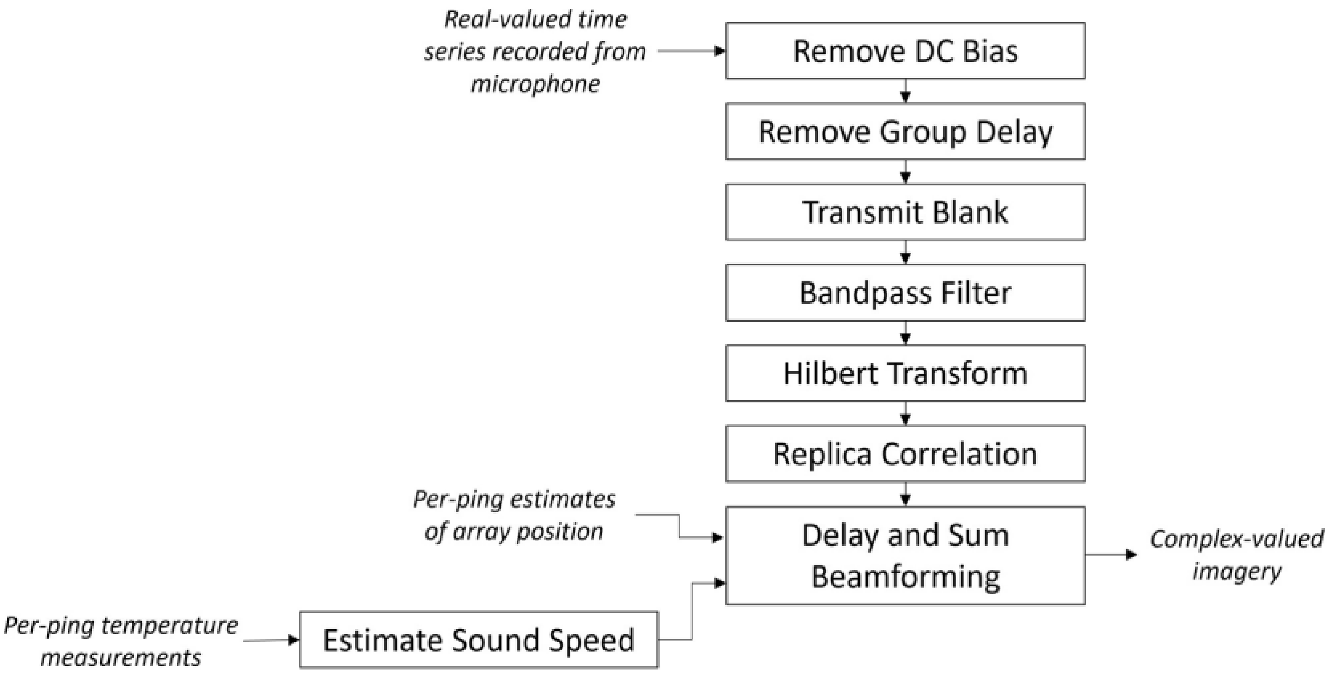

Figure 3.8: Raw data processing flow
[15].

Figure 3.8 displays the different steps of data processing, from raw data to the processed signal ready to be used for feature extraction. The processing pipeline is implemented in the MATLAB code supplied alongside the dataset [71]; this code was adapted for the present work. These steps are also useful to minimize non acoustic artifacts. In this section, all the different steps involved will be explained in detail and why each one of them is useful for the final classification. The raw data from all the experiments that were collected by the sonar is stored in a database as time series, meaning that the signal is in the time domain, with time in the x axis and amplitude of the collected signal in the y axis.

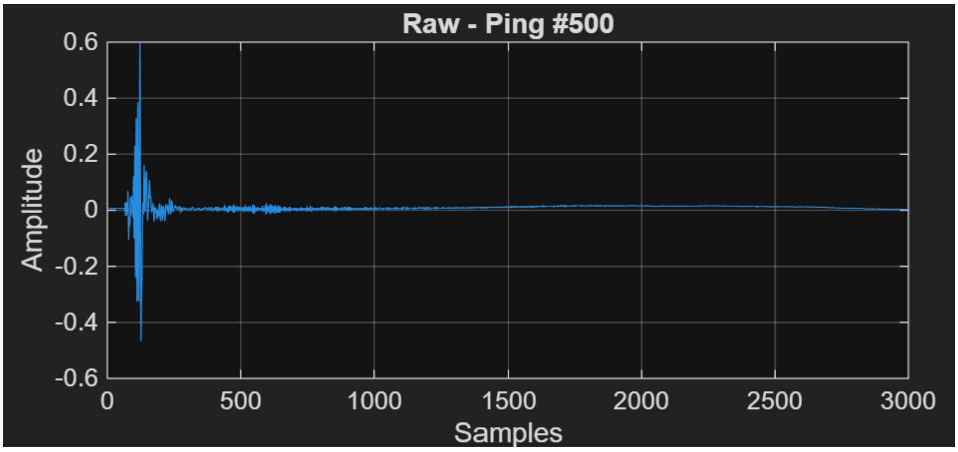

Figure 3.9: Time series of the raw data. Graph created in MATLAB with the code provided by Blanford et al. (2024)
[15]

Figure 3.9 shows the time series of the raw data, from ping number 500 for the letter Q targets, partially buried in the rough surface. The x axis is expressed in samples but it can easily be converted in time, as the sample rate is given and its value is 100.000 Hz [15]. The ping contains information from the echoes of whole scanned scene and thus from the targets under interest. Although for this time series, data is still too raw to be used for classification. The echoes are barely visible to not visible at all and the initial burst of energy from the sonar completely overpowers the rest of the echo. The different steps of the data processing flow will be illustrated with the graphs of the pings, to show how the signal is changed for every step of the process.

3.2.1.1 DC bias removal

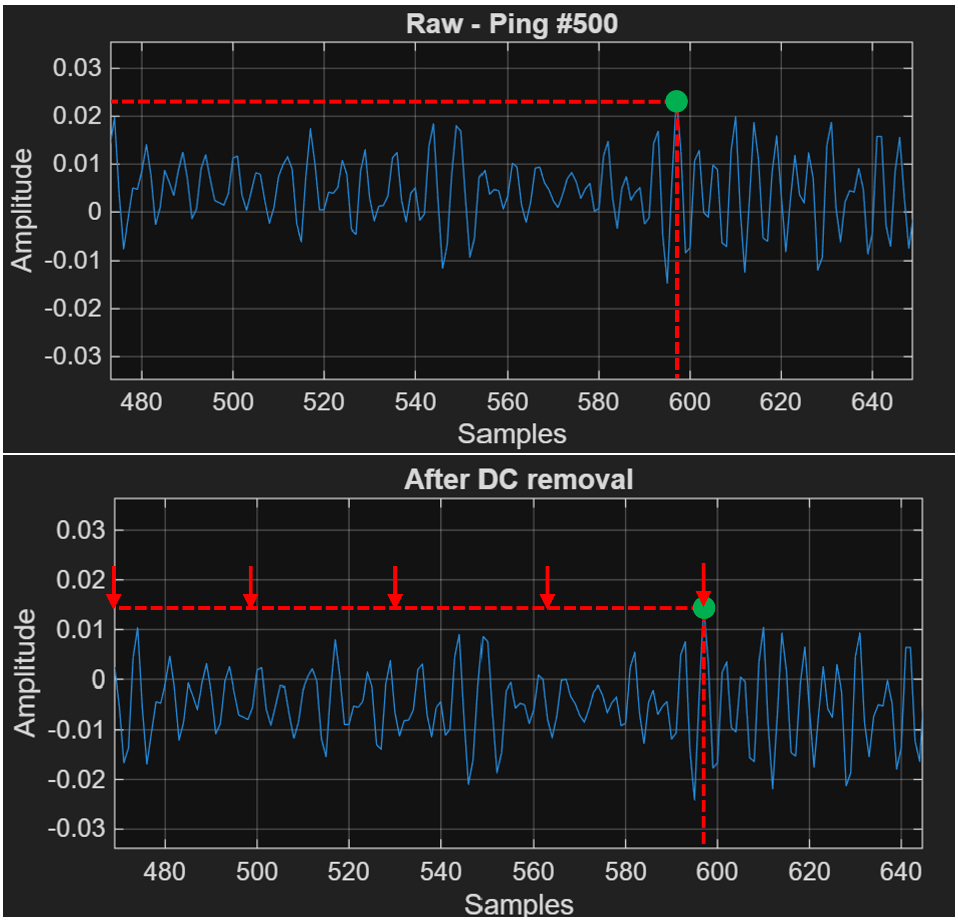

Figure 3.10: Zoomed in time series from raw data (left) and after DC removal (right) for comparison purpose. Arrows show the direction of the shift. The green point shows the position of the same data (arbitrarily chosen) before and after shifting. Graphs created in MATLAB with the code provided by Blanford et al. (2024)
[15]

The first processing step shown in figure 3.8 is the DC bias removal. The main cause of DC bias in an acoustic setup the Analog-to-Digital Converter (ADC). The role of an ADC is to convert a continuous analog signal, which is the pressure variations of the echoes, into a digital signal that can be read by a computer. It basically samples the signal in several points. The sample rate tells how many points there will be, and the higher its value, the more precise the conversion will be. But this hardware is the cause of an offset in the signal, shifting the whole signal down and the compensation of this offset ensures that the mean of the entire signal is centered at zero on the y-axis. Furthermore, this bias is constant throughout the whole signal. Figure 3.10 displays the difference of offsetting the DC bias back to the natural position of the true signal with respect to the raw data [72]. The whole signal is, in fact, shifted down. If this step is overlooked, the next processed from the data processing flow may be compromised, as the mean of the signal needs to be around zero for these processes to be applied properly.

3.2.1.2 Group delay removal

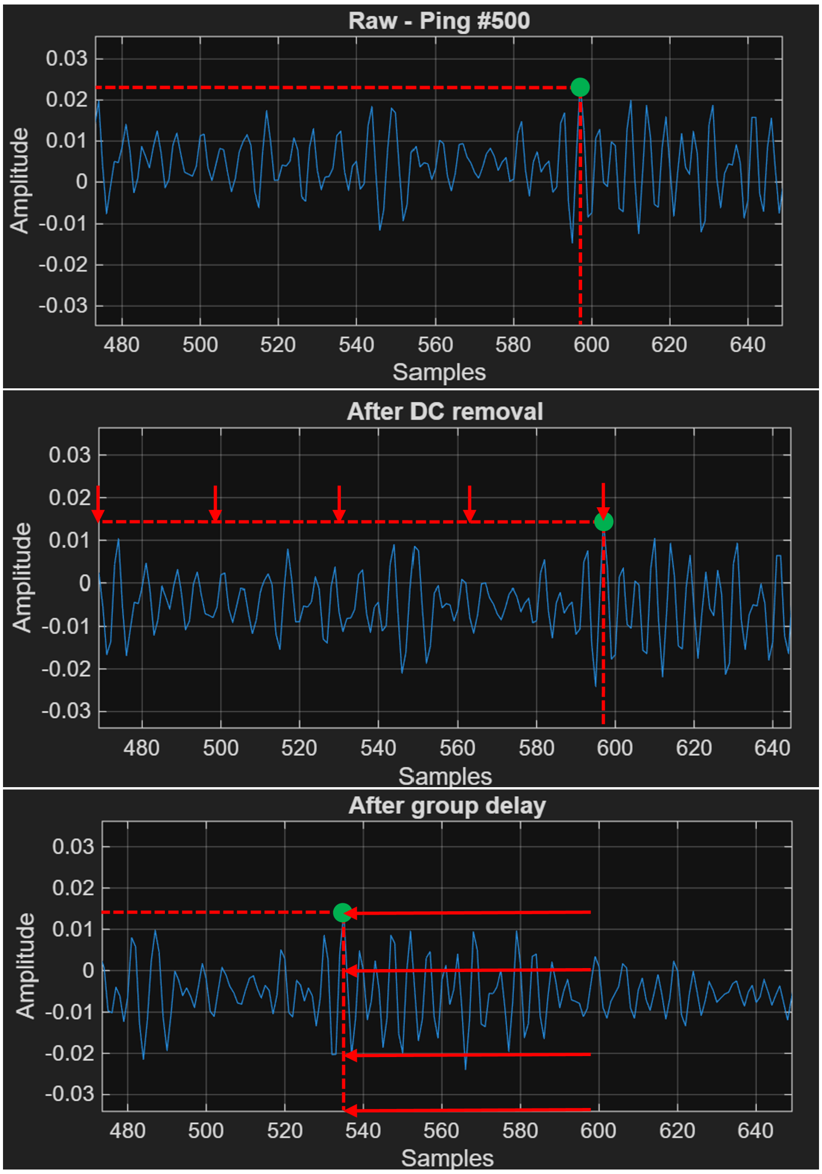

Figure 3.11: Zoom in the time series from raw data (left) after DC removal (center) and group delay removal (right) for comparison purpose. Arrows show the direction of the shift that is induced by the DC removal (central figure) and by the group delay (bottom figure). The green point shows the position of the same data (arbitrarily chosen) before and after shifting. Graphs created in MATLAB with the code provided by Blanford et al. (2024) [15]

The second processing step involved, as seen in figure 3.8, is the removal of the group delay. Its main goal is to cancel the offset caused by the hardware of the acquisition system [73]. There are several factors that can cause such a delay. The first factor is that the signal travels through the different electrical wires and compounds of the acoustic system, and even though the speed of the signal is fast, it still takes a non-negligible amount of time. So, one has to account for this time that causes a slight delay in the data. Another factor is linked with electronic latency and the last factor is caused by a possible imperfect synchronization between the transmitter and receiver of the sonar. The compensation of the delay works involves a transformation of the time series to the frequency domain for the delay compensations and an application of the delay, which is known after being calculated in a calibration step beforehand. The applied shift is not visible to the naked eye in the whole ping. Therefore, a zoomed version is presented, in figure 3.11, along with a zoom in the graphs of the raw data and after DC removed. The whole signal is shifted to the left. The marked point represents the same point in the echo and was shifted for around 60 samples. The shift to the left makes sense because the delay causes the whole data to appear later in time than it should, so shifted to the right in this cartesian representation with the origin of the axes at the left.

3.2.1.3 Transmit blanker

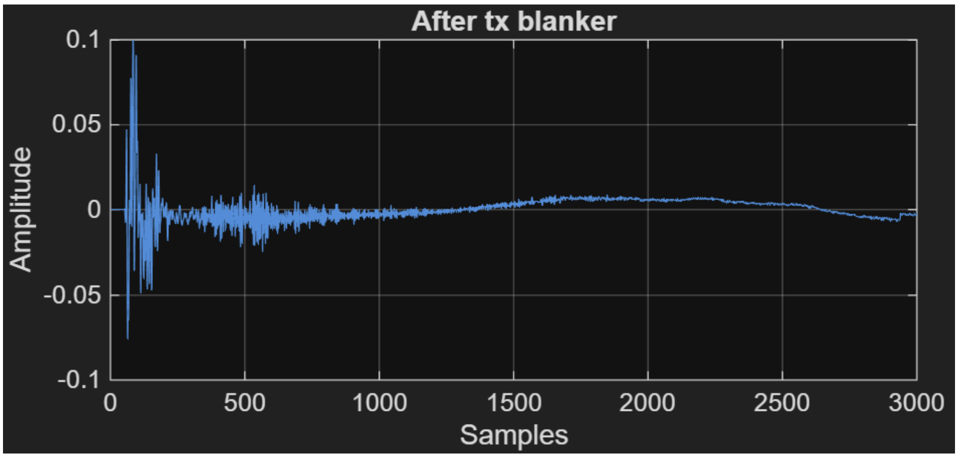

Figure 3.12: Graph of the ping after the transmit blank. Graph created in MATLAB with the code provided by Blanford et al. (2024) [15]

In this step, the main goal is to remove the contribution of the incident wave in the time series, because the power of the transmitted wave is far greater than that of the reflection and transmission waves [74]; moreover, the burst is emitted directly next to the microphone, this is why it appears at the beginning of the ping. The blank basically suppresses the inputted chirp signal and only leaves the echoes after the emission in the time series, in other words removing the signal that travels directly from the transmitter to the receiver. Because, in the raw data, the first burst completely overshadows the rest of the signal and this is exactly what this step prevents. On figure 3.12, it is clearly visible and the rest of the signal becomes more readable and exploitable. The maximum value went from 0.6 to 0.1. Furthermore, characteristic parts of the signal begin to appear. The signals coming from the targets closer to the sonar at the moment of the ping appear first in the time series, and with more power so with a greater amplitude. The blanker works by calculating the duration of the input chirp signal, in sample units, and suppresses, in the signal, the values for this amount of samples. A second step involves a smooth transition to get a progressive increase in the beginning of the signal.

3.2.1.4 Bandpass filter

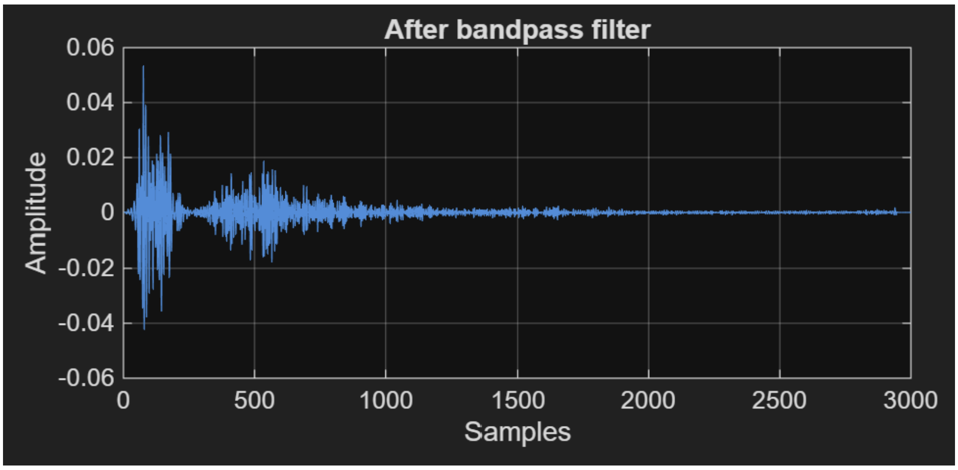

Figure 3.13: Graph of the ping after the bandpass filter. Graph CREATED IN matlab with the code provided by Blanford et al. (2024)
[15]

A bandpass filter is a tool that removes some parts of a signal that are out of a given frequency band. The filter used in this data processing is a special kind of filter called a high-pass filter. It allows higher frequencies to remain intact while mitigating lower frequencies. The reason for using such a filter is that the incident signal is a chirp that sweeps a range of frequencies, and frequencies found below a certain level are unwanted low frequency noise caused by vibrations. This filter set the threshold (or cutoff) frequency to 8 kHz, the lowest frequency value swept by the LFM (10 kHz) reduced by 2 kHz to get a smoother transition from the suppressed signal to signal that is kept. In short, a high-pass filter is mainly used to get rid of unwanted noise and improve Signal-to-Noise Ratio (SNR). It is also a prerequisite for the next processing step described below. The global signal was presented an oscillation, that was staightened by the high-pass filter as seen on figure 3.13. This oscillation was caused by a slow drift of the amplitude, as seen from around 1000 to 3000 samples in figure 3.12. This drift is the result of noise caused by the nominal rate of oscillation of alternating current. It is possible to estimate the value of the slow drift, which can be calculated with simple mathematics:

$$f = \frac{1}{T} \tag{3.9}$$

$$T = \frac{N}{f_s} \tag{3.10}$$

Equation 3.9 shows the link between frequency and period, while equation 3.10 displays how to calculate the period with the number of samples N and the Sampling rate f_s. On figure 3.12, after a visual analysis, the number of samples in the half oscillation is around 1000 samples, The whole oscillation measures approximately 2000 samples, give or take 200 samples. at a sample rate of 100 kHz, the period measures 0.02 seconds. The frequency of the slow drift is $1/0.02$ = roughly 50 Hz, which corresponds to the noise generated by current in the grid in Europe (the value of this frequency depends on the voltage of the grid and is therefore not the same everywhere on earth [75]) and is called the mains frequency [76]. This noise reverberates on the measured data because of the simple fact that all hardware is connected to the grid.

3.2.1.5 Hilbert transform and analytical signal

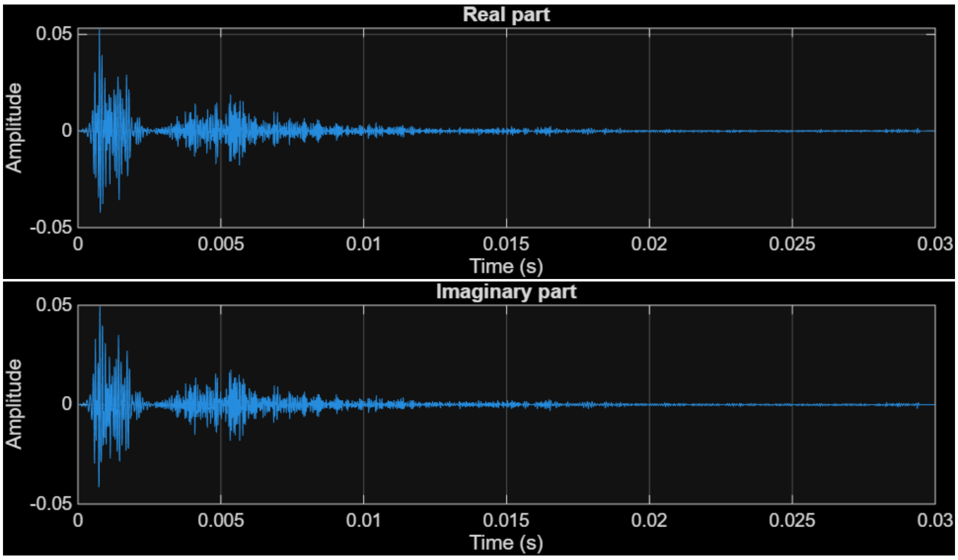

Figure 3.14: Real (upper image) and imaginary (lower image) parts of the signal. Graphs created in MATLAB with the code provided by Blanford et al. (2024) [15]

A Hilbert transform is a tool that is used to build an analytical signal from a real valued signal. An analytical signal is a complex valued signal containing the real signal and its conjugated Hilbert transform. This type of signal is also called the envelope of the signal . The Hilbert transform works by shifting the phase of a signal by 90°, without touching the amplitude [77]. A simple way to represent this is that the Hilbert transform of a sine function is the cosine function. The two separated signals of figure 3.14 are identical on every aspect but the phase, which is shifted, so visually they practically look alike. But the interesting part is the reconstructed envelope containing both signals as seen in equation 3.11.

$$z(t) = x(t) + j.\mathcal{H}\{x(t)\} \tag{3.11}$$

$$|z(t)| = \sqrt{x(t)^2 + \mathcal{H}\{x(t)\}^2} \tag{3.12}$$

$z(t)$ is called an analytic signal. The application of a Hilbert transform $\mathcal{H}\{x(t)\}$ produces two important features that were hidden in the real signal:

- The first feature is the envelope $|z(t)|$. This is the module of the signal constructed with the original signal and its Hilbert transform And is the amplitude given at every instant. Another characteristic of the envelope is that it only contains positive values of the signal, as it is obtained by calculating the absolute value of the real signal and its imaginary counterpart as seen in 3.12 [78];

- The second feature is the instantaneous phase. $arg(z(t))$. Its derivative gives the instantaneous frequency. It tells at which frequency the signal is oscillating at every instant.

The real valued signal is like a projection, with the global information contained within one dimension. By calculating the envelope, the signal is now a two-dimensional signal, making hidden properties appear, allowing to gain access to the whole potential of the data. The envelope tells the amount of energy that was reflected by the targets, and thus the intensity of the reflections is measurable, regardless of the frequency of the signal, making it a powerful tool for data analysis. Figure 3.15 represents the envelope. Because the absolute value was taken, only positive values appear.

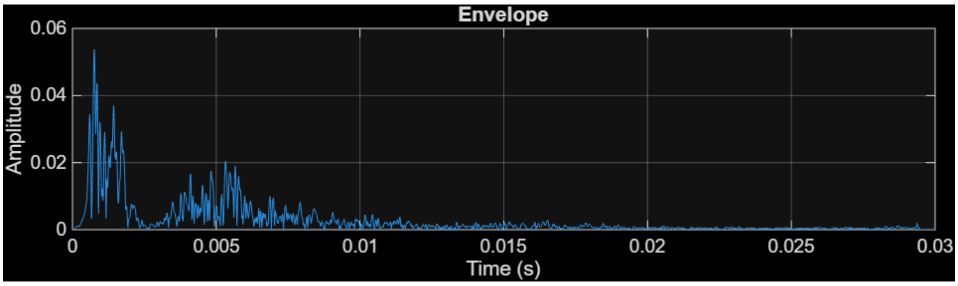

Figure 3.15: Graph of the absolute value of the envelope of the complex valued signal. Graph created in MATLAB with the code provided by Blanford et al. (2024) [15]

3.2.1.6 Matched filter

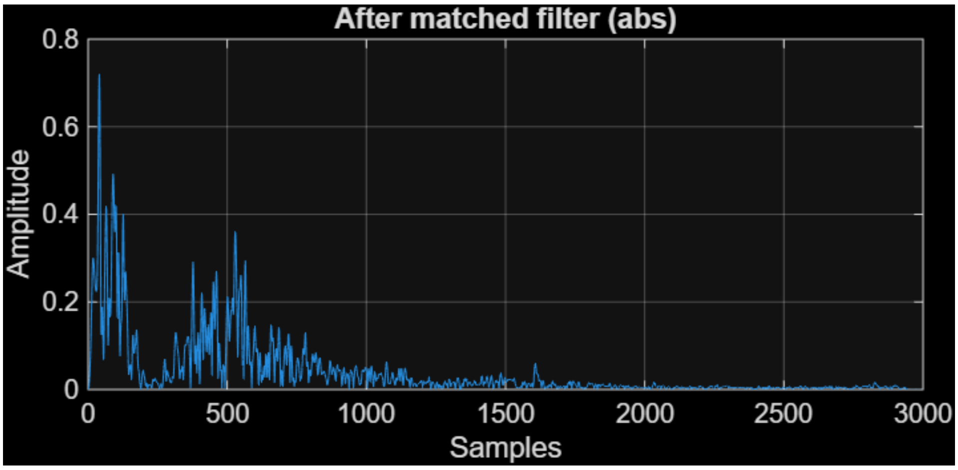

Figure 3.16: Graph of the signal after replica correlation. Graph created in MATLAB with the code provided by Blanford et al. (2024) [15]

A matched filter, also called a replica correlation, is an operation in which the collected signal is compared, by correlation, to the known incident signal, called the template. In this case, the template is the chirp, and the main goal is to detect its presence in the collected signal or, in other words, to find where the chirp contributions are located within the response signal. So, the template is glided through the whole time series and similarities are calculated by correlation. This is beneficial by maximizing the SNR. Another definition for match filters is that they are the filters that yield the highest SNR [79]. There are a variety of matched filters, and the one used here is a frequency matched filter. The first step of the matched filter involves a Fast Fourier Transform (FFT), which transforms a time domain signal into a frequency domain signal. The second step is the correlation itself as shown in equation 3.13.

$$Y(f) = X(f).H^*(f) \tag{3.13}$$

X(f) is the signal after FFT, $H^*(f)$ is the complex conjugate, inverse of the template chirp signal, and Y(f) is the new correlated signal. The reason behind the phase inversion of the template is to make sure that in the correlation, when a spike is aligned with the template, the correlation is positive and the signal is amplified. For the rest of the signal, weak correlations make sure that the uninteresting parts stay low in the time series. After that, an Inverse Fast Fourier Transform (IFFT) is applied to get back to the time domain. Looking at the graph of the signal after pulse compression in figure 3.16, some parts of the signal have gained some amplitude, with the maximal amplitude reaching almost 0.8, for 0.6 in the previous step. It shows that the correlation by pulse compression enhanced the parts of the signal that correspond to the reflection of the targets. Some of the previous steps of the processing pipeline were necessary for the matched filter to work properly. A signal containing the prior bias from the DC offset (as explained in subsection 3.2.1.1) or the bandpass filter (explained in subsection 3.2.1.4) would contain too many external artifacts that were caused by signal pollution from the hardware and other electrical vibrations. The application of the pulse compression would lead to false correlations with unintended signal features, and the final signal would be much more different from what it was intended to be.

This was the last step involved in the processing flow. The step delay and sum beamforming shown in figure 3.8 relates to SAS processing, which is not used here, for reasons already explained in section 2.5. The data is now ready to be used for classification, but there is one last step before proceeding. The useful data first has to be extracted from the global data.

3.2.2 Data extraction

Data extraction is a key step for classification. Many algorithms need to be fed data of a specific kind, and the algorithms used for this study use features, which are quantities that give information on the signals. The algorithms do not work with actual time series. Therefore, the challenge is to extract the features that represent the signals in the best possible way. The extraction of the data explained in this section contains three different steps, comparable to three layers of data extraction, with each layer diving deeper into the data than the previous one. The first step is the extraction of useful pings from the overall set of thousands of available pings. The second step involves the selection of useful echoes within the pings themselves, because within a whole time series, only some parts are actually interesting for the classification. The third extraction step aims to compute features from useful echoes. Those features are the very last step towards classification.

3.2.2.1 Extraction of the pings

As explained in section 3.1, the experiments created several thousands of time series, each containing information on the targets. But, the sonar's transmitter has a limited FOV of 120°. The scene in which the targets are placed is quite large, and some of the pings do not contain information of the furthest targets, because the data contained within the FOV contain around 80% of the energy of the signal. For these pings, if the data is extracted at the location where the targets are supposed to be, there will mainly be noise, as shown in figure 3.17, for target 6 and 7. The process of signal windows extraction will be explained in the next subsection.

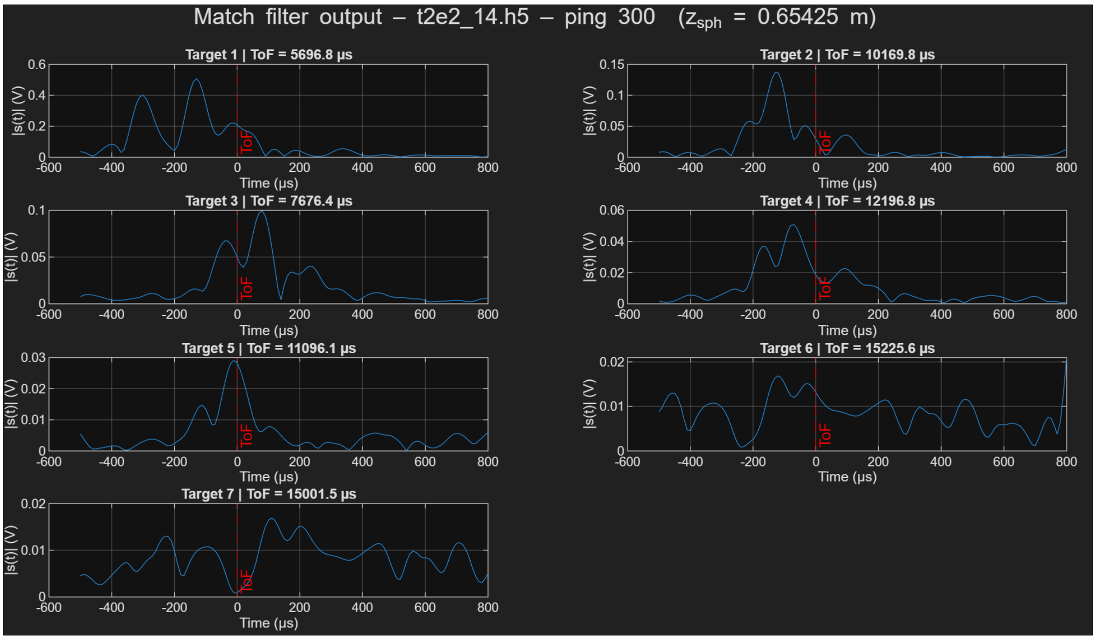

Figure 3.17: extracted windows of signals for targets of ping number 300 from hollow aluminum sphere on flat interface, collection 14. Graphs created in MATLAB with the code provided by Blanford et al. (2024)
[15]

Ping number 300 is included in the first third of the scene as there are 1000 pings for one scan. Therefore it would be expected to have proper signals of the first targets (namely targets one to five maximum, as seen on figure 3.3 from section 3.1.1). Figure 3.17 shows this behavior as the closest targets exhibit the highest energy levels and the clearest reflection signals. They show nice spikes of high energy and internal reflection echoes after the specular reflection because the target is elastic. The energy of the targets gradually decreases as the targets are located further from the receiver of the sonar, with target one the closest, which has the shortest time of flight of the acoustic wave, and the highest energy. The signals of targets six and seven are mainly background noise because the energy of the reflections of the targets was at the same level as the noise, which shows an amplitude varying under 0.02.

Due to this phenomenon, there were two possibilities to extract the data. The first was to select the useful signals within every ping. That would mean calculating, for every ping, all targets that were contained within the range of the sonar. The other possibility was to calculate a range of pings that would contain useful data from all the targets and apply this first filter for all the scans of all the experiments. This was also decided so that all the target signals would therefore have, more or less, similar amplitudes. After a little bit of calculations, the 'good pings' are those ranging from ping number 476 to ping number 625. As the targets were always placed at the same locations in the virtual boxes (see figure 3.3), the selection of pings would always lead to the extraction of all the data from all the targets and it would constitute enough data for the dataset even though some of the data were not used.

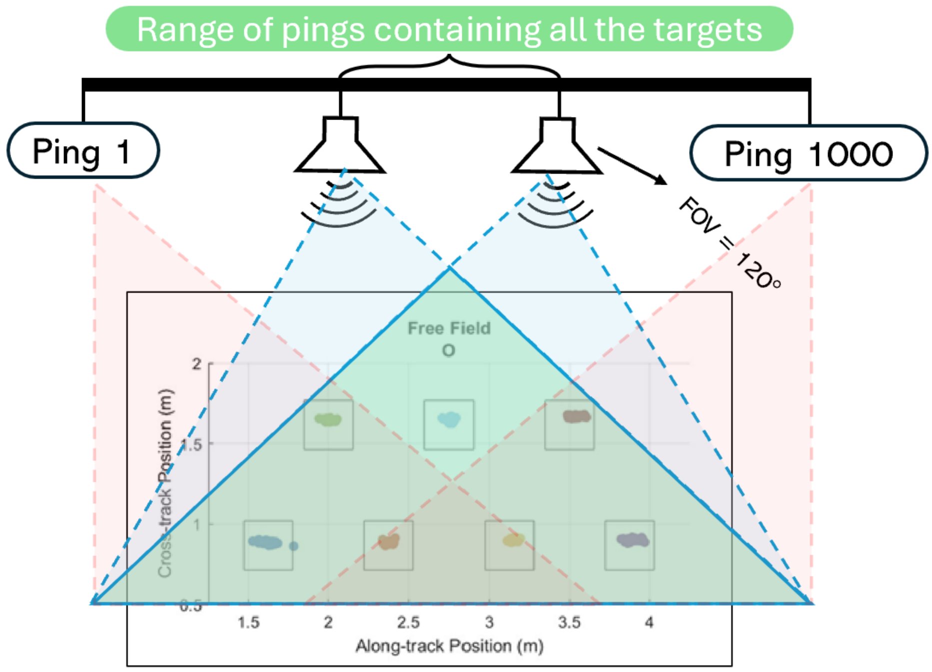

Figure 3.18: Illustration of the range of pings that were selected. The image was self-made.

This figure (3.18 shows the motivation behind the ping selection. If pings located at the very end of the scene were kept, they would fail to capture data from some of the targets. the red triangles correspond to the FOV of the sonar for the first and the last pings, and it is clearly visible that some of the targets are out of reach. The blue triangles represent the FOV of the sonar for the pings in the filtering range. They cover all the targets (green overlapping in the figure) and should therefore be kept. This figure is for illustration purpose only and is not at scale.

In addition to this range selection, a second filter was applied based on spatial proximity. The spatial difference between two pings is five millimeters. Therefore, the difference in angular shift between them is quasi negligible. In terms of classification, keeping echoes that have so many similarities would mean a lot of data repetition. This would cause data leakage, overfitting, and biasing in the classification. This is why it was decided to keep only one out of ten pings in the range that was previously established. In other words, one ping every five centimeters, which would give enough angular variation for the echoes to be significantly different. So, for one scan, fifteen pings were kept, between ping 476 and ping 625.

To summarize this step, two filters were applied to the ping extraction. The first one is the choice of a range of useful pings, containing clear and thorough echoes from all the targets. the second one is to maintain only one ping out of ten, to prevent similar echoes from appearing.

3.2.2.2 Extraction of the windows

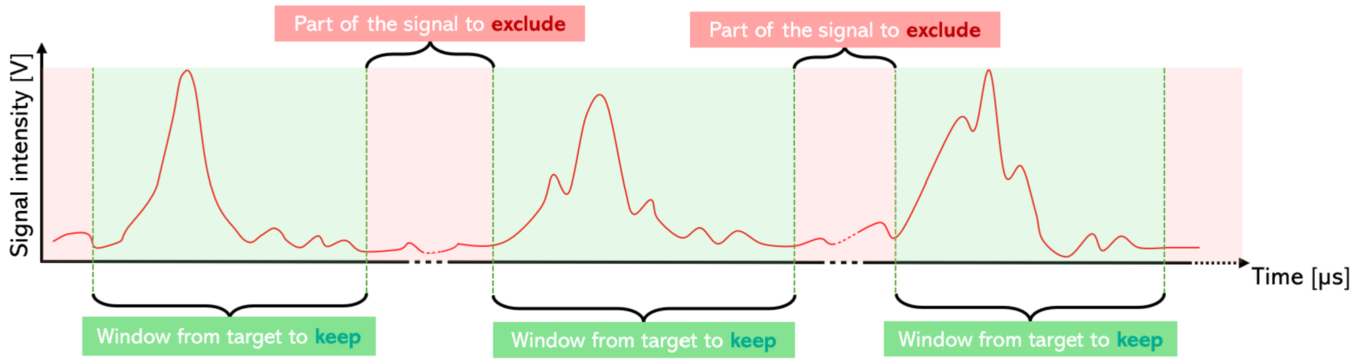

Figure 3.19: Illustration of windows extraction for one ping. The green areas are those that are kept. The red areas are excluded. The image was self-made.

Now that the right pings have been chosen, it is time to dive within the pings themselves. A ping is the time series corresponding to the global echo heard by the receiver of the sonar. For every take, the microphones were listening for 0.03 seconds. At a sampling rate of 100.000 Hz, one ping contains 3000 samples. So the ping is like one long function with a lot of unwanted data mainly containing noise from the environment. The actual, useful data of the targets therefore have to be extracted. Fortunately, the exact positions of all the targets are known. At least the virtual boxes in which they are randomly placed (see figure 3.3). The time of flight of every virtual box could then be calculated, and their coordinates were chosen as the center of every box The coordinates of these targets are shown in table 3.2. The positions given for Tx and Rx are respectively the initial positions of the sonar's transmitter and receiver. Actually, 4 microphones were used for this experiment. The one chose here was the closest one, to be as close as possible to a monostatic sonar. These positions were updated from 5 millimeters to the next position for every ping. The distances traveled by the sound wave were calculated with the following equations:

$$d_{Tx-target} = \sqrt{(X_{target} - X_{Tx})^2 + (Y_{target} - Y_{Tx})^2 + (Z_{target} - Z_{Tx})^2} \tag{3.14}$$

$$d_{target-Rx} = \sqrt{(X_{Rx} - X_{target})^2 + (Y_{Rx} - Y_{target})^2 + (Z_{Rx} - Z_{target})^2} \tag{3.15}$$

$$ToF = d_{Tx-target} + d_{target-Rx} \tag{3.16}$$

With all the known positions, it was possible to calculate, for every ping, the ToF of the sound wave for every targets in all concerned pings. This ToF was then used to find the exact location of the seven echoes within the pings. A window around that ToF was then taken. And to do so, the size of the virtual boxes in which the targets were placed had to be taken into account. They measured approximately 30 cm by 30 cm. this size has to be converted in seconds with the speed of sound, which roughly corresponds to 500 μs. The window also has to be asymmetric to take into account the targets that were potentially placed at the far right of the boxes. In such cases, the echo linked with internal vibrations, which lasts for a certain amount of time (which depends on the composition of the targets, because the speed of sound is not the same for MDF, aluminum or polyurethane), would be out of the window if it was symmetric. So on the left of the signal the window was set at 500 μs and on the right, it was set at 800 μs to take this possibility onto account.

Table 3.2: Positions of the targets and the transducers. The data comes from Blanford et al. (2024) [15]

Target	X position [m]	Y position [m]	Z position [m]
1	1.625	0.866	0.654
2	2.000	1.616	0.654
3	2.375	0.866	0.654
4	2.750	1.616	0.654
5	3.125	0.866	0.654
6	3.500	1.616	0.654
7	3.875	0.866	0.654
Tx	0.000	0.000	1.094
Rx	-0.045	-0.0013	1.094

But there was still a remaining problem. Because the data are contained in the time domain, targets that were located at the same distance from the sonar, even if they were at a different location, would have the same time of flight. That means that in the time series their echo would overlap, which makes their usefulness basically vanish. The features extracted from such signals would have information on both targets, potentially multiplying the energy and ultimately biasing the calculated features.

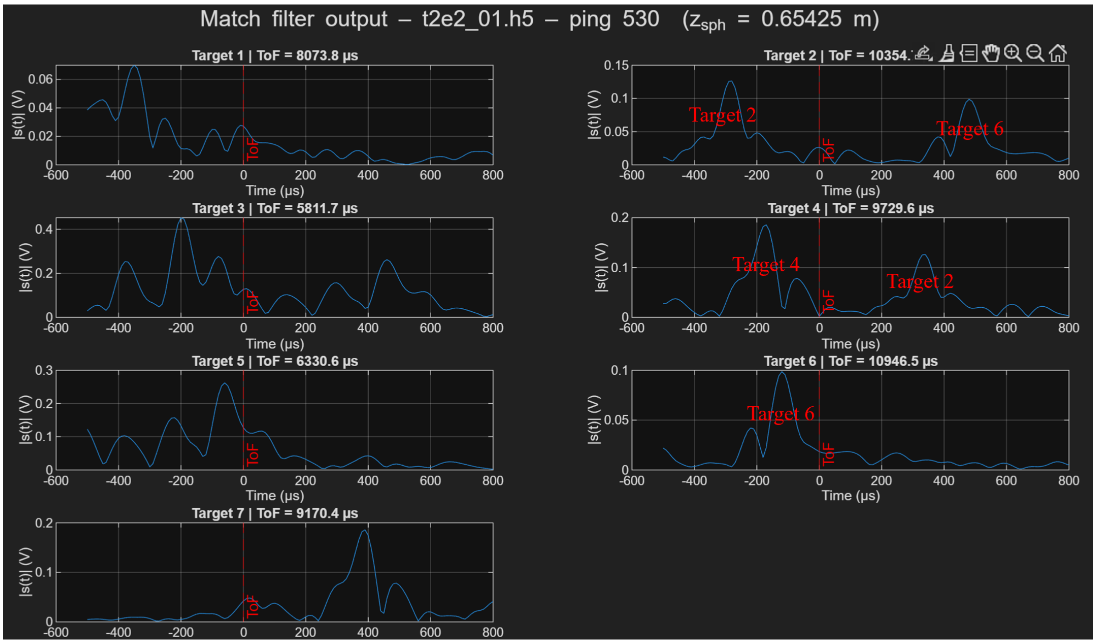

Figure 3.20: extracted windows of signals for targets of ping number 530 from hollow aluminum spheres on flat interface, take number 1. Graphs created in MATLAB with the code provided by Blanford et al. (2024) [15]

The time series in figure 3.20 is particularly speaking. Targets 2, 4, and 6 have a very similar time of flight and, therefore, are located at a similar distance to the sonar. Their signals clearly contain information from each other. For instance, the graph of target 2 also contains the echo of target 6 and r-the graph of target 4 contains the echo of target 2. To overcome this problem, a secnd filter has to be applied to the data. The time of flight of all the targets were compared with each other, and when the value was too close, both echoes were abandoned to avoid the overlapping problem. A threshold was needed and was arbitrarily chosen to be 700 μs. This value was chosen because it had to be not too low to avoid overlapping but also not too high to avoid losing information on the echoes of the targets. Therefore, 700 μs seemed to be a good compromise after several tests. In this case, targets 6 and 2 would be excluded, but target 4 too, because its time of flight is too close with that of target 7 by a small margin, though. Targets 3 and five would be excluded, too, for the same reason, leaving only target 1 to be kept.

At first glance, it may seem that these extraction steps, for the pings and within the pings are too restrictive, and would take a lot of data away. While this is true, there is so much available data that the extracted data is enough to build a separate dataset. The ping extraction leaves around 2 to 3 echoes of targets per selected ping and 15 pings were selected for every scan, which means that per scan there are roughly 40 windows selected. Now, by looking at figure 3.4, where the number of scans are displayed for every target and every environment, it is possible to calculate the amount of target windows available for the whole dataset. It is important to note that the scans involving no targets were not used. The summation of all used scans gives 552 scans, multiplied by 40 target windows, the dataset contains 22,080 windows. This amount of data is enough for machine learning classification.

To summarize the steps of this extraction process, within the pings themselves. The first step was to extract windows in the whole time series to get the desired echoes of all the available targets. The second step involved the exclusion of windows containing information of several targets, because their ToF were too similar.

3.2.2.3 Extraction of the features

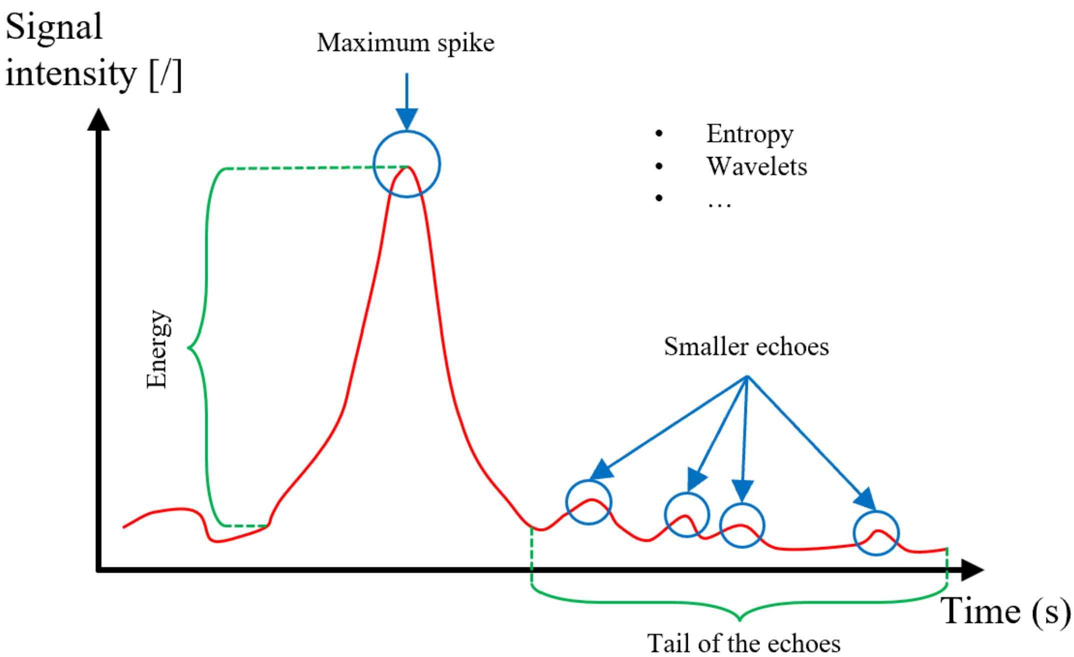

Figure 3.21: Illustration of some features extracted by tsfresh. The features are characteristics found within the echo, that distinguish all echoes from each other. The image was self-made.

The data is almost ready to be fed to the algorithms. The last step is to extract features from the echoes. A feature can be seen as a value giving information on the signal, such as the number of spikes, the energy, the maximum and minimum values, and so on as seen in figure 3.21. The features were extracted for the envelope of the signal, but also for real and imaginary signals.

The tool used to extract those features is tsfresh, a Python tool that automatically recognizes and calculates hundreds of features. The data fed to tsfresh were, at first, the envelope of the signal and both real and imaginary components. The results were hundreds of features that could finally be used in the Machine learning classifiers.

However, many of the features are useless because they are not correlated in any way to the targets' physical properties, and their implementation would slow the calculations down without adding any benefit. Even worse, these unnecessary features could cause bias in classification. So, the goal is to maintain the features that are really necessary for this purpose.

The selection of features was made by calculating the correlation between all the features and the rigidity or elasticity class extracted in the data labels. of the targets. The correlated features were compared between the different classes, and a cross-selection was made to take the features that are present for all types of environments and targets. The resulting number of features is less than 200, without affecting the classification power.

The correlation method used is the Spearman's rank Correlation method, also called Spearman's ρ. This correlation method measures the link between two variables but not based on their values. Rather, it measures the correlation on the basis of their rank. The coefficient ranks between ρ = -1 (negative correlation) and ρ +1 (positive correlation) and if $\rho = 0$ in the middle for an non-existing relationship between the values. The first step in the correlation assessment is to sort the parameters in an ascending order, then giving each parameter its rank in the order. This method allows to rank the features from the most correlated to the one which is the least correlated with the properties of the targets. The Spearman's ρ is used here to measure the extent to which a feature is linked to the target's material, by looking how the feature changes if the target's material type changes. If the value of a feature stays the same in both cases, it would mean that it is not linked with the material type. Such features are the ones that are most likely linked with background noise, or features present in both rigid and elastic classes.

The advantage of using this method is to measure a monotonous relation between the target's material and a given feature by looking at the order of that feature. Furthermore, this correlation method is robust to outlier values. A method to assess the robustness of a correlation is to look at the p-value. If the p-value of a correlation is lower than a threshold value (generally 0.05 or, in this case, 0.01), it is considered significant.

3.2.3 Classification methods

In this section, the different classification algorithms used to classify targets on their elasticity or rigidity will be presented. This part is crucial and the parameters of the algorithms have to be fine tuned to the data to improve classification performances. Three different algorithms were used for classification purposes, all categorized as supervised machine learning algorithms but can be separated into two categories. The first category encompasses the algorithms that use decision trees to classify. This type of classifiers is called non linear. The second category contains regression algorithms that are not based on decision trees but instead calculate a linear combination of the variables These algorithms are, in contrary to the latter, linear algorithms and thus, of simpler structure.

All classification algorithms share the same general working principle: the available data are split into two sets. The first set is called the training set and contains most of the data, usually around 80%. This data will serve, as its name intends, to train the algorithm to fit the data as well as possible. The algorithm hyperparameters must be selected carefully to ensure proper training. The remaining part of the data, so usually 20% of the whole set is called the test set, and is used for the classification itself. The performances of a classifier mostly rely on the quantity of data in the dataset, and to get interesting results the test set also has to be sufficient.

As stated earlier, the algorithms used here are called supervised. Actually, machine learning algorithms come in two main classes: Supervised and unsupervised [80]. Unsupervised, as shown in figure 3.22, functions as follows: the algorithms group the data in clusters by using automated methods or algorithms on data that was not categorized or labeled. Such algorithms must learn relationships and features by seeking patterns within the data and build different groups containing data sharing similarities. On the contrary, supervised learning algorithms (as depicted in figure 3.23) generalize knowledge from data that has been labeled and classified to train on and then to be able to make predictions on unlabeled data with the constructed algorithm. Figures 3.22 and 3.23 were self made, using PowerPoint.

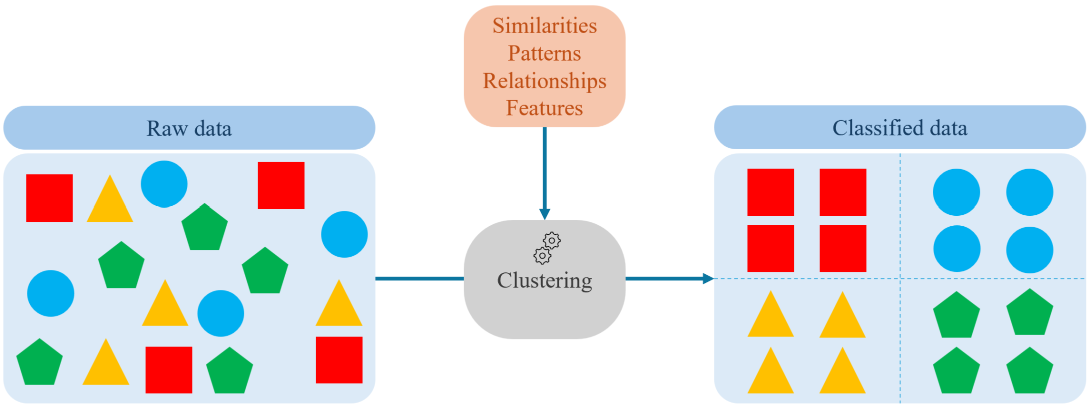

Figure 3.22: Illustration of the functioning of unsupervised machine learning. This illustration was self-made.

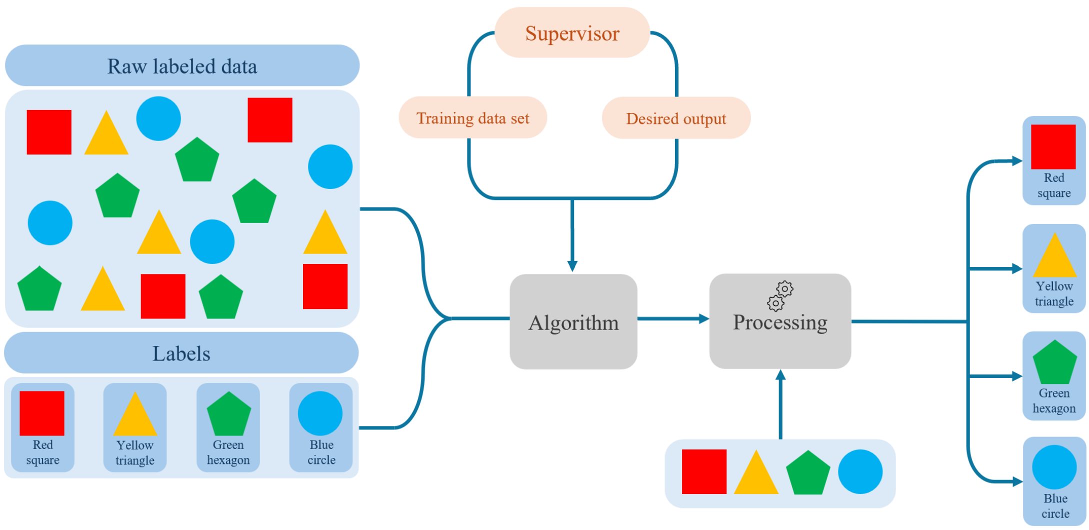

Figure 3.23: Illustration of the functioning of supervised machine learning. This illustration was self-made.

The core functioning of a decision tree relies on three distinctive steps. The first step is the algorithm inducing the growing and the pruning of the tree's functionalities. The second step is the growth of the tree. The third and last step consists in the pruning of the fully grown tree. This step is necessary to optimize the models performances.

A key challenge to classification algorithms is overfitting. This term is often used in the literature when it comes to classification and with machine learning algorithms, in general. Overfitting is defined as when a model fits too closely or even exactly to its training data [81]. The problem with that is that when overfitting happens, the model isn't able to make predictions on other datasets, because it only knows the training data so well that it fails to understand anything that differs to it even by a slight margin. This defeats the purpose of machine learning as the goal is to be able to make predictions of unknown data based on the training of known data. How does this happen ? When the algorithm trains too long on the known data, it starts to earn the noise characteristic to the training data. Therefore, when a classification is being made, it is necessary take this phenomenon into account and take precautions to avoid it.

3.2.3.1 Random forest

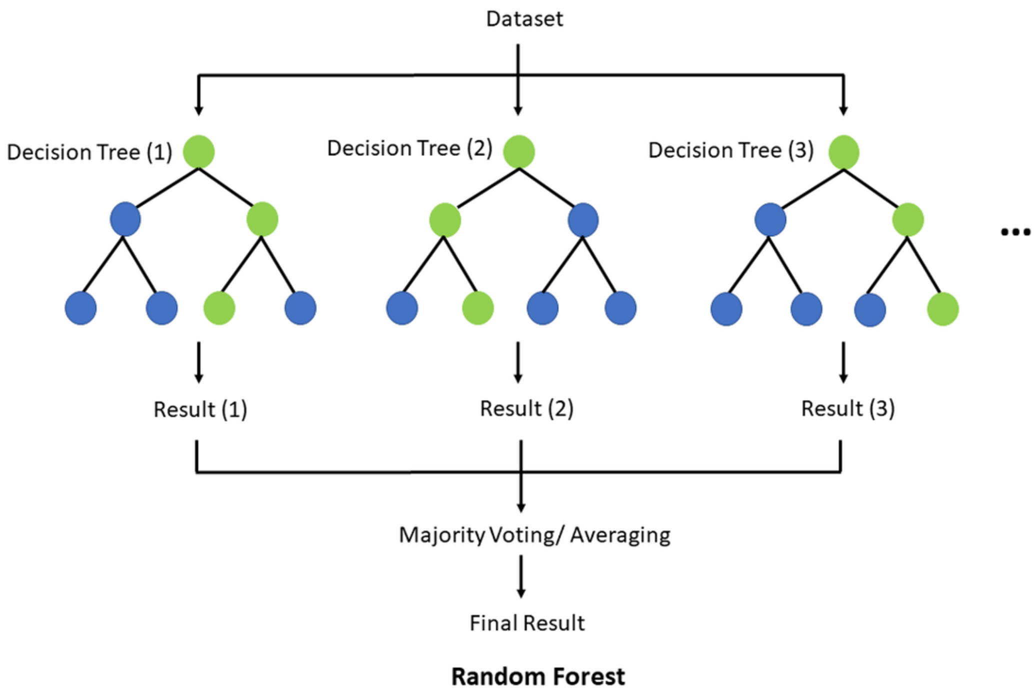

Figure 3.24: Illustration of a random forest algorithm. Several trees are created simultaneously, and for each tree, several branches separate from decision nodes, leading to new nodes. The results of each tree are then used to calculate a global result for the random forest. The green dots are the pathways decided by the choices made at each note, while the blue dots are the choices that were not kept.
[82]

The random forest is one of the most commonly used classification algorithms due to its ease of use, flexibility, and powerful performance. It can be used for classification or regression purposes.

A random forest [83] functions by generating a multitude of decision trees at the same time, hence the name random 'forest'. A decision tree can be seen as a flow chart of conditions that acts as a means to split the data. Every condition is a node that divides in two branches. The node checks if the data verifies the condition or not and the answer will lead the data towards one of the next 2 nodes, accordingly to the answer to the condition. Each data value is assigned to a class, based on the value of the target variable that is the most represented at the moment of iteration. Basically, at each node, a choice is made as depicted here in a simple example, where x is the value of data fed to the classifier and 5 is a threshold value at a given node:

$$\text{if } x < 5 \;\rightarrow\; \text{go to the left branch}$$

$$\text{otherwise} \;\rightarrow\; \text{go to the right branch}$$

All trees are built together at the same time, and the outcome of each individual tree is averaged to calculate the overall performance of the classification.

Random forest algorithms are practical and their benefits are numerous:

- Reducing overfitting. It overcomes the risk of overfitting linked with decision trees as they tend to tightly fit the samples within the training data, thanks to the use of many trees at the same time. The average of uncorrelated trees lowers the overall variance and prediction errors;

- Easy determination of important features. There exist several tools that assess how performances decrease when some variables are removed, providing an overview on the important data and, moreover, the data that can be excluded.
- This algorithm is also efficient at dealing with datasets in which there are missing data.

But there are also drawbacks to these algorithms [84] [85]:

- Less reliable with unbalanced datasets;
- High computation costs. Because the creation of a multitude of trees simultaneously requires a lot of computing power. This is why reducing the amount of features is important;
- Slow previsions. Because high computational power is required, Real time applications are severely reduced;
- Reduced interpretability compared to single trees. Random forests, which build several trees simultaneously, act as a black box;
- Overfitting on noisy data.

3.2.3.2 XGBoost

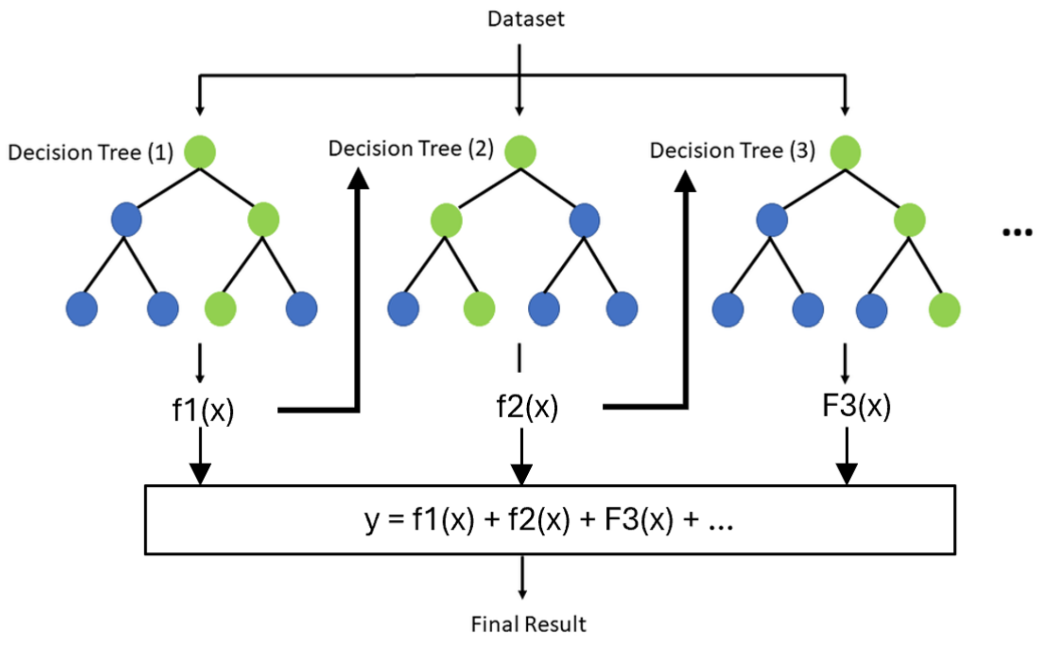

Figure 3.25: Illustration of the XGBoost algorithm. The green dots are the pathways decided by the choices made at each note, while the blue dots are the choices that were not kept. Each residue from a decision tree (here fi(x)) is used in the calculation of the next decision tree. [82]

XGBoost is another example of supervised machine learning classification algorithm. but it differs from the random forest in the sense that, instead of computing several trees at the same time, it generates one tree at a time [86]. It works by generating a first decision tree with the data from the training set. a residual is obtained from this first tree and is used in the calculation of the next tree. This new tree aims to improve the previous tree and correct its errors thanks to the use of its residue. All the other trees are built by means of the same logic, so that each new tree is an improvement

of the previous one as depicted in figure 3.25. The final result is the sum of the residues of each tree, with each residue being smaller than the previous one. This is where the name of the algorithm comes from (eXtreme Gradient Boosting), the whole algorithm is a boosted gradient from the first tree to the last.
The benefits of the XGBoost algorithm are also numerous. XGBoost is:

- Fast. This algorithm is created in such a way that it allows parallel computing for training. There is integration of computing features that allow to speed up the processes [86];
- Effective at limiting overfitting. It integrates regulators that penalize complex models and reduces overfitting [86];
- Able to work with smaller datasets [87] as well as big dataset [86];
- Able to compensate for missing values in the dataset [88];
- Flexible, such as the random forest algorithms, working for classification as well as regression problems [86].

3.2.3.3 Logistic regression

The last type of machine learning algorithms that has been used to perform classification is logistic regression. This type of classifier is a supervised algorithm, like the random forest and XGBoost classifiers. But unlike the latter two, it does not rely on decision trees to make predictions. Logistic classifiers rather make prediction as the probability of an input to belong to one of the classes and is frequently used in binary classification. The logistic regression has been widely used across many domains, including sonar target classification for its high analytical power and adaptability in determining the relationship between two or more variables [89].

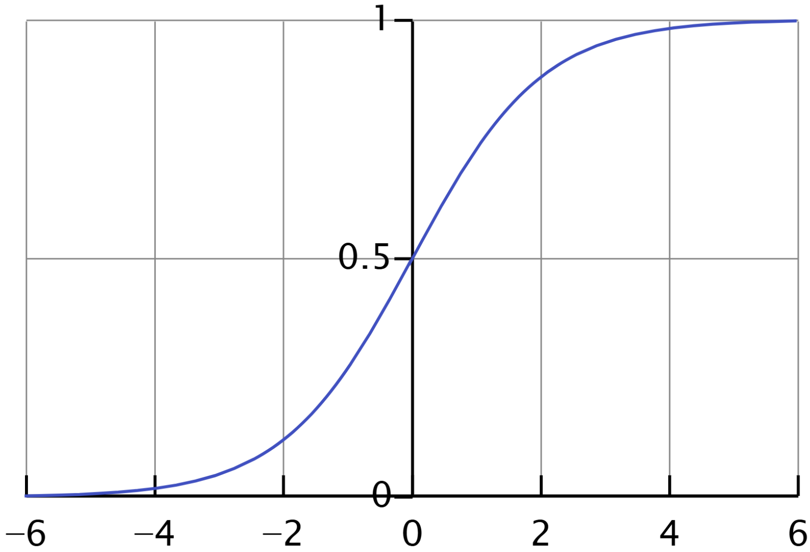

Figure 3.26: Typical sigmoid function

At their core, lies the use of sigmoid functions [23], rather than linear functions that are used in linear regression calculations. In linear regression, the algorithm tries to find the line that best fits the data by calculating the sum of least squares. A sigmoid curve has an "S" shape, see figure 3.26, The general equation is depicted in equation 3.17, which is used to categorize the input data into a probability value between 0 and 1, following the sigmoid curve. A threshold value is chosen, which is usually 0.5 to be right in the middle of the probability distribution, and every data that is below this threshold is categorized as 0 and all the data above 0.5 is placed in the 1 category.

$$S_x = \frac{1}{1 + e^{-x}} \tag{3.17}$$

The logistic regression algorithm comes with its share of benefits and down-points. First, let us explain the down-points [90]:

- The first down-point comes to the fact that the prediction is in the form of a probability and not a single valued prediction. Therefore, interpretability is complex. In contrast to the other classifiers that yield a yes or no, this one gives the result as a spectrum of probabilities. Therefore, the choice of the threshold directly influences the classification, which reduces interpretability;

- Difficulty in the establishment of the significant variables from all the input variables, which can lead to over-saturation of the model if insignificant variables are kept.

Now let us dive in the benefits of using logistic regression:

- Ease of use and low computing power required;

- Well suited for binary classification

The results of all these classification algorithms can be evaluated with a confusion matrix containing explanatory values such as the recall, the accuracy or the prevalence (measure of the actual elastic targets in the overall target population). Furthermore, metrics such as the ROC-AUC curves can serve as a strong analysis tool to evaluate the performance of a classifier [90].

Chapter 4

Results

4.1 Confusion matrix and the performance metrics

As explained in the section classification, classification was performed on all four separated environments, as well as for all the environments together, to evaluate the performances of the classifiers. The classification results showed great variation from environment to environment. The two less complex environments yielded the best results among all the classifications. When plastic pellets were added to the scenes, performances dropped significantly. The results were the lowest when the targets were partially buried in the pellets. Classification was also performed with features that were automatically pre-filtered by the tsfresh algorithm, as well as the more advanced feature filtering using correlation and cross-selection, as explained in section 3.2.2.3, to assess if the performances could be either at least the same or enhanced with the feature selection.

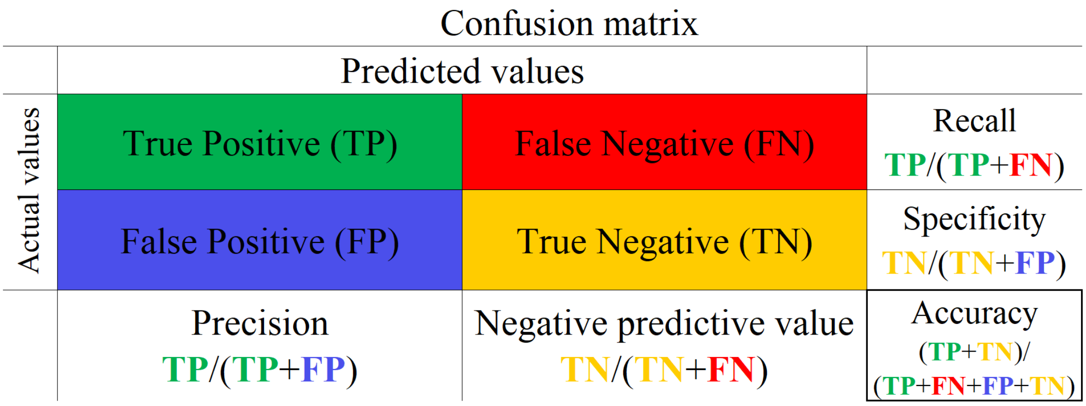

Figure 4.1: Structure of a confusion matrix.

Figure 4.1 depicts the structure of a confusion matrix and the most relevant metrics that can be calculated with the classification results [91]. The matrix is divided into rows representing the actual values of the data and columns representing the prediction values of the classifiers. For a binary classification, the matrix simply consists of two lines and two columns. The true positive (TP) and true negative (TN) values correspond to the targets that were correctly placed in the right category. In this case, TP corresponds to true elastic targets and TN corresponds to true rigid targets. The two other values, false negative (FN, also called the type II error) and False positive (FP, also called the type I error) are all the targets that were misplaced in a wrong category. So, the FN category corresponds to all elastic targets that were placed in the rigid category, while FP contains the rigid targets that were placed in the elastic category. These values are then used to calculate a series of metrics that give information on the classification capacities of the algorithms that are easier to interpret than simply looking at values comprised in the matrix itself. A perfect classifier has a

100% score in all metrics, because the FP and FN values would be equal to zero. The accuracy, for instance, tells the overall performance of the classifier, and is therefore very helpful. The higher this metric, the better the classification. Precision is the fraction of relevant instances among the retrieved instances and recall is the fraction of relevant instances that were retrieved.

4.2 Feature selection results

Tsfresh first calculates 794 features, according to the documentation available online and, with extraction performed on the three available signals, the real part, the imaginary part, and the reconstructed envelope, that accounts for a total of 2382 features. this is a large number to feed the classification algorithms and a large portion of these features could be irrelevant, because not linked with the targets' materials and shapes. The correlation and cross-selection steps of the feature filtering yielded 197 features that were all most relevant features that were present in all environments.

4.3 Results with global features

Tsfresh integrates its own feature selection, with efficient parameters and a relevance filter. According to the description available in the documentation of the tsfresh python plugin, the number of features selected depends on the data, and thus it is not a constant value among all the environments. But in general, the number varied around several hundreds. These features are labeled global features, as the number of features is greater with the filter integrated within tsfresh than the number of features extracted manually by correlation and cross-selection. These features, on the contrary, are called the filtered features for comparison purposes.

Table 4.1: Classification results for flat interface environment, with random forest (a) and XGBoost (b) and with global features.

(a) Random forest, free-field

	Predicted Elastic	**Predicted Rigid**
True Elastic	507	20
True Rigid	5	303

Metric	**Value**
Recall	0.96
Precision	0.99
Accuracy	0.97

(b) XGBoost, free-field

	Predicted Elastic	**Predicted Rigid**
True Elastic	513	14
True Rigid	2	306

Metric	**Value**
Recall	0.97
Precision	1.0
Accuracy	0.98

The results in table 4.1 were almost perfect for the free-field environment, with great recall, precision and accuracy, with similar results for both algorithms. XGBoost beats the random forest in accuracy with a very slight margin. These results highlight the capability of the classifiers in low noise level conditions.

Table 4.2: Classification results for flat interface environment, with random forest (a) and XGBoost (b) and with global features.

(a) Random forest, flat interface

	Predicted Elastic	Predicted Rigid
True Elastic	467	16
True Rigid	5	583

Metric	Value
Recall	0.97
Precision	0.99
Accuracy	0.98

(b) XGBoost, flat interface

	Predicted Elastic	Predicted Rigid
True Elastic	471	9
True Rigid	2	584

Metric	Value
Recall	0.98
Precision	0.99
Accuracy	0.99

For the flat interface environment, the results in table 4.2 were even better than for free-field, by a small margin. This may be explained by the fact that background noise can originate from the entire room in free-field conditions and the targets are suspended by fishing wires, which may introduce minor disturbances.

Table 4.3: Classification results for rough interface, proud with random forest and XGBoost respectively (a and b) and partially buried with random forest and XGBoost respectively (c and d) and with global features.

(a) Random forest, rough interface proud

	Predicted Elastic	Predicted Rigid
True Elastic	69	171
True Rigid	12	296

Metric	Value
Recall	0.29
Precision	0.85
Accuracy	0.67

(b) XGBoost, rough interface proud

	Predicted Elastic	Predicted Rigid
True Elastic	69	171
True Rigid	18	290

Metric	Value
Recall	0.29
Precision	0.79
Accuracy	0.65

(c) Random forest, rough interface partially buried

	Predicted Elastic	Predicted Rigid
True Elastic	234	216
True Rigid	73	263

Metric	Value
Recall	0.52
Precision	0.76
Accuracy	0.63

(d) XGBoost, rough interface partially buried

	Predicted Elastic	Predicted Rigid
True Elastic	252	198
True Rigid	87	249

Metric	Value
Recall	0.56
Precision	0.74
Accuracy	0.64

The classification results for the two more complex environments are lower than for the two previous conditions. But the most relevant information to keep in mind is the notably low recall and precision values, indicating that the classifier assigned most targets to the rigid category. In the matrix, it is seen in the true negative (TN) cell, which contains most of the classified targets in almost all cases. Another interesting characteristic to note is that the recall is constantly lower for the proud interface than for the partially buried interface. However, this is probably caused by the imbalance in the data availability for the two types of targets. For the targets that were laid proud, 240 windows were extracted for elastic targets and 308 from rigid targets, while for the partially buried targets, 450 elastic target windows were extracted for 336 rigid target windows. Once again, the performances of both algorithms are very similar.

4.4 Results with filtered features

Table 4.4: Classification results for random forest free field (a) XGBoost free-field (b), random forest flat interface (c) and XGBoost flat interface (d).

(a) Random forest, free-field

	Predicted Elastic	Predicted Rigid
True Elastic	502	25
True Rigid	8	300

Metric	Value
Recall	0.95
Precision	0.98
Accuracy	0.96

(b) XGBoost, free-field

	Predicted Elastic	Predicted Rigid
True Elastic	510	17
True Rigid	11	297

Metric	Value
Recall	0.97
Precision	0.98
Accuracy	0.97

(c) Random forest, flat interface

	Predicted Elastic	Predicted Rigid
True Elastic	467	13
True Rigid	4	584

Metric	Value
Recall	0.97
Precision	0.99
Accuracy	0.98

(d) XGBoost, flat interface

	Predicted Elastic	Predicted Rigid
True Elastic	466	14
True Rigid	3	585

Metric	Value
Recall	0.97
Precision	0.99
Accuracy	0.98

Overall, the results with far less, but meticulously selected features yielded strong results in table 4.4, with similar results as with global features. In some cases, the results were even slightly better. like the accuracy for the random forest classification for flat interface, which is slightly above the classification performed with global features. This shows that a careful selection of features enables to perform classification without losing classification power and even in some cases to improve the classification.

Table 4.5: Classification results for Rough interface environments, proud (a and b) and partially buried (c and d) with random forest and XGBoost and filtered features.

(a) Random forest, rough interface proud

	Predicted Elastic	Predicted Rigid
True Elastic	62	17
True Rigid	13	295

Metric	Value
Recall	0.26
Precision	0.74
Accuracy	0.65

(b) XGBoost, rough interface proud

	Predicted Elastic	Predicted Rigid
True Elastic	51	189
True Rigid	12	296

Metric	Value
Recall	0.21
Precision	0.81
Accuracy	0.63

(c) Random forest, rough interface partially buried

	Predicted Elastic	Predicted Rigid
True Elastic	232	218
True Rigid	80	256

Metric	Value
Recall	0.52
Precision	0.74
Accuracy	0.62

(d) XGBoost, rough interface partially buried

	Predicted Elastic	Predicted Rigid
True Elastic	252	198
True Rigid	87	249

Metric	Value
Recall	0.56
Precision	0.74
Accuracy	0.64

The results of classification in table 4.5 with filtered features for the more complex environments are also very similar to the ones from global features, with the same conclusions to be taken for free field and flat interface; the filtering of the features yields similar results, with far less input data and thus less computing power is needed.

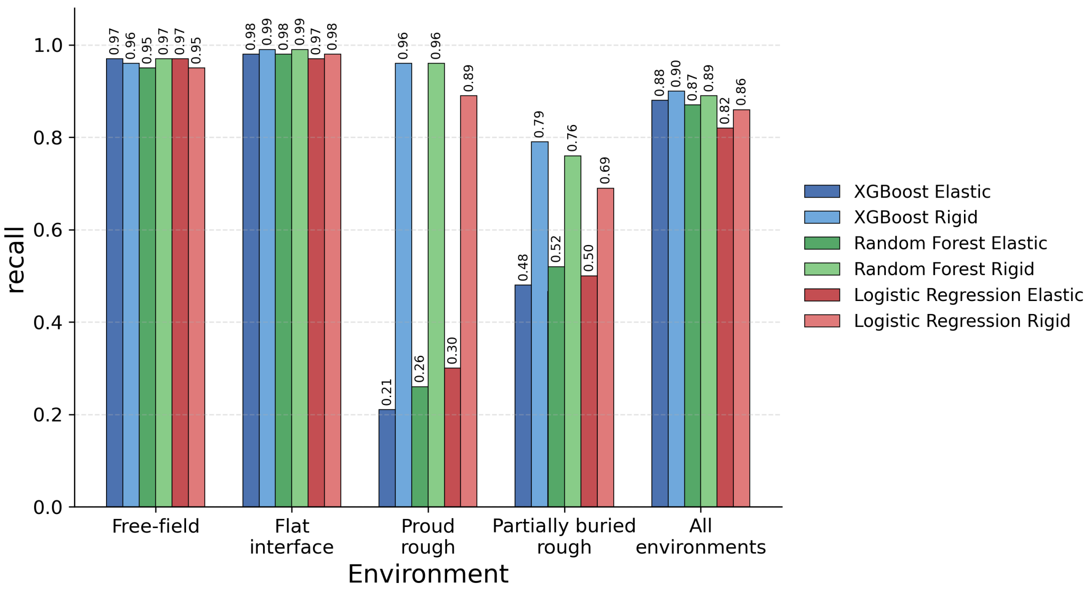

Figure 4.2: comparison of recall values for all environments.

This figure shows the recall for all classifiers and all environments under study. These values are almost always very good for the rigid class, even when the elastic class is bad. This shows a strong bias of the data towards the rigid class, meaning that something is altering the data from the complex environments. This could be caused by the plastic pellets, making up the rough interface. The pellets could actually scatter the acoustic waves in a way that the echoes of the elastic targets contain signatures that are similar to the rigid class, signatures that are generated by the pellets.

4.5 Classification of all environments together

In real life conditions, the targets are not all present in one of the four environmental conditions, but rather they are found in all of them simultaneously. Some of them will be buried or partially buried, whereas others will be buoyant in the water. This is why it was decided to perform classification on a dataset containing data from all four environments, as displayed in figure 4.2.

Table 4.6: Classification results for all environments at once with random forest (a) and XGBoost (b) and with filtered features.

(a) Random forest, all environments

	Predicted Elastic	Predicted Rigid
True Elastic	702	102
True Rigid	82	664

Metric	Value
Recall	0.87
Precision	0.89
Accuracy	0.88

(b) XGBoost, all environments

	Predicted Elastic	Predicted Rigid
True Elastic	707	97
True Rigid	77	669

Metric	Value
Recall	0.88
Precision	0.90
Accuracy	0.89

The classification results obtained in table 4.6 are quite good, reaching an accuracy of almost 90%. More importantly, the recall and precision values are good, the classifiers are therefore able to distinguish classes far better than for the complex environments alone. Again, performances of both algorithms are very similar, with a variation of less than 1%.

4.6 ROC-AUC curves

As explained in the beginning of this chapter, there are several metrics that are helpful in the analysis of the results of a classification algorithm. The metrics calculated with the values of the confusion matrix are a good estimation of the overall performances but there exists another method of representing the performances. This method is called the ROC-AUC curve. It helps to understand how well the classifier distinguishes between the rigid and elastic classes [91].

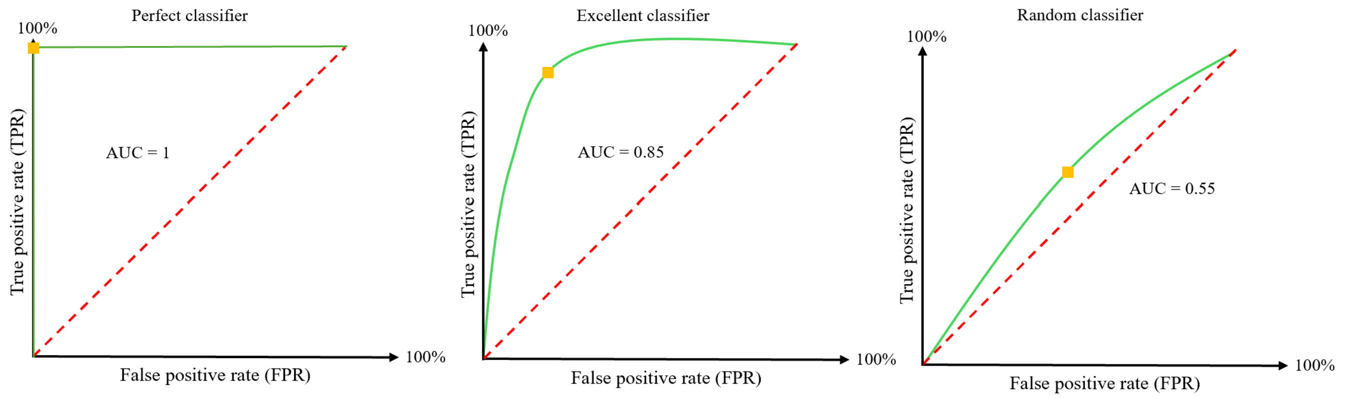

Figure 4.3: ROC-AUC curves of binary classifiers from perfect to random

The power of the ROC-AUC curve, as shown in figure 4.3, is that, contrary to the confusion matrices that only show results for one threshold (usually 0.5, as for this study), the ROC-AUC curve plots the True Positive Rate (TPR) and the probability of False Positive Rate (FPR) for all threshold values at once. The threshold is defined as such: During classification, the model does not directly assign a label to each observation, it produces a probability score between 0 and 1, and it is the decision threshold that sets the boundary from which a score is interpreted as positive (i.e., elastic) in this classification. The better the classifier, the more the curve will approach the upper left corner of the plot. A perfect classifier has an AUC equal to one. This translates the fact that a perfect classifier has a TPR of 1 and an FPR rate of 0. That means no classification errors, as seen in figure 4.3. on the right hand plot. A random classifier will have an AUC of around 0.5 as shown in the right hand plot of figure 4.3. A ROC-AUC curve is represented for one of the two classification classes. For instance, if the material type chosen to be represented is rigid, the y axis will contain the true rigid rate and the x axis will contain the false rigid rate. However, True Negative Rate (TNR) - the true rate for the other class - and FPR are linked: TNR=1-FPR, so both curves are the same, but one is the mirror image of the other around the axis of the random line (for an AUC=0.5).

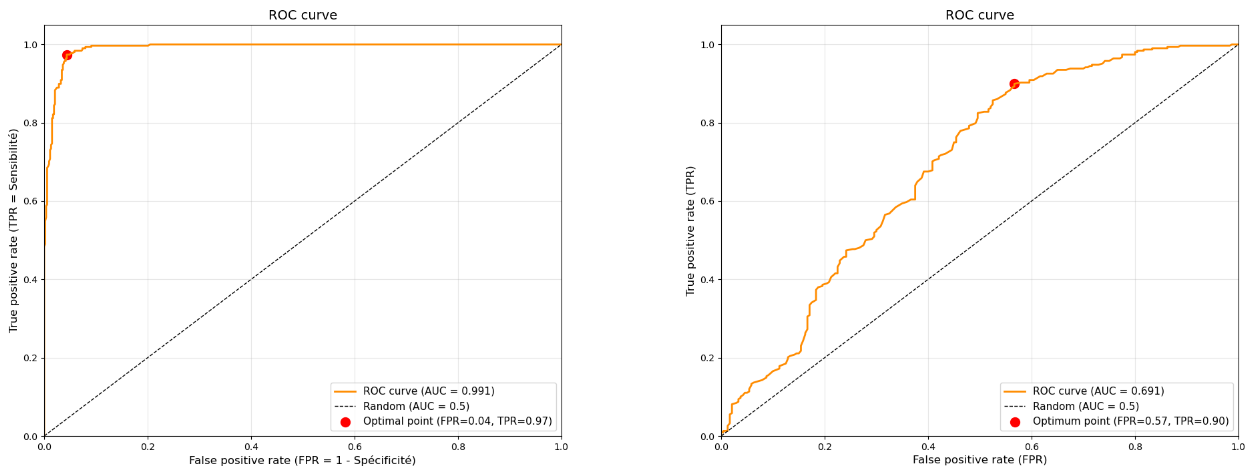

Figure 4.4: ROC-AUC curves of random forest classification in free field (left) and rough interface proud environments, filtered features, for the rigid class.

The curves depicted in figure 4.4 are two of the most extreme cases and a good comparison point for the analysis of this type of curves. On the left, the AUC is almost at its highest possible value and resembles the left plot in figure 4.3. while the ROC curve from the classification performed with the targets proud on rough interface is much closer to the random line, as expected. The direct visual analysis of a ROC-AUC curve is an intuitive way of obtaining a first assessment of on the capabilities of the classifier and the curve on the right is clearly too shifted to the right.

4.7 F1 scores

The F1 score is another powerful tool that is used to assess the performance of a classifier [91]. What makes it so useful is that it accounts for false positives and false negatives at the same time, as seen in equation 4.1. It is defined as the harmonic mean of the recall and the precision. This value can be calculated for both classification classes and is appropriate for unbalanced classifications when a class is more present than the other, which happens in this case. The fact that this metric integrates both recall and precision is also a limitation in its interpretability and does not take the true negatives into account, this is why it is important to include the score for both classes.

$$F1 = 2 \times \frac{precision \times recall}{precision + recall} = \frac{2TP}{2TP + FP + FN} \tag{4.1}$$

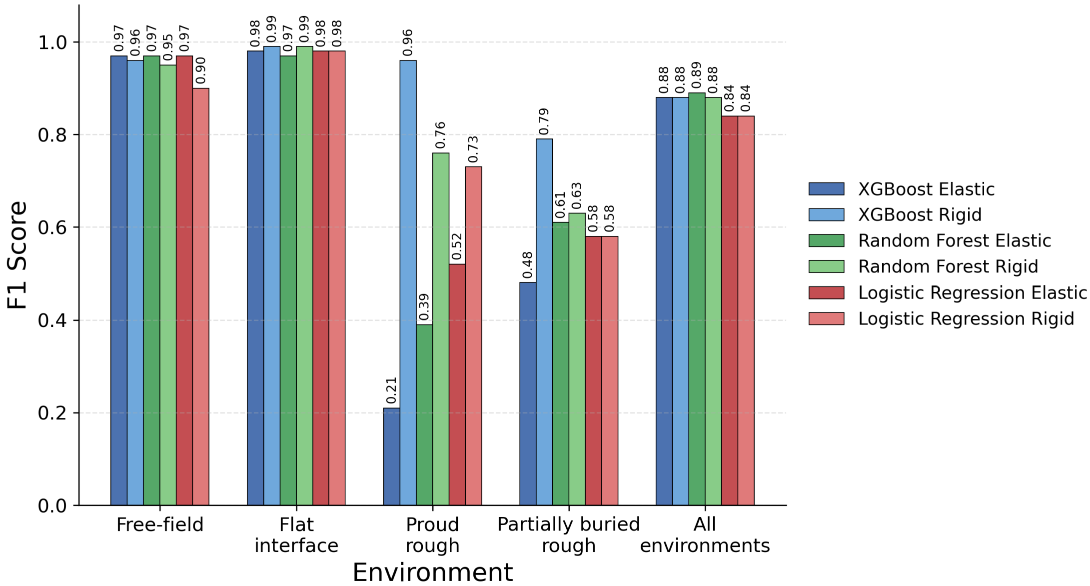

Figure 4.5: F1 scores for all environments and for both classes

The representation of the various F1 scores visually shows the classification capabilities of the different classifiers. For the rough interface, XGBoost tends to be considerably better for the rigid targets than the elastic ones. Actually, for the proud targets on rough interface, the performances regarding rigid targets are almost as high as the ones for less complex environments. The problem is that the classifiers are biased toward this class and the other results are way too poor, like the F1 score of the elastic class for the proud on rough interface environment being really bad.

4.8 Subtraction of the blanks

The fact that the results were poorer for the complex environments led to the hypothesis that the cause of the problem could come from the plastic pellets making up the rough interface, and that their reflections may cause artifacts that are interpreted as rigid by the classifiers. therefore, as an attempt to improve classification in the rough interface environment, it was decided to subtract the background scans, which correspond to recordings of the scene without targets, but in the same conditions as if there were targets. In figure 3.4, the presence of such blanks is shown. The mean value of all blank scenes was subtracted from the raw acoustic data with targets, proud or partially buried in the rough interface. Then, the data processing pipeline and the ping and target windows extractions were applied, as usual, to build a new dataset. This was done to remove all potential noise generated by the pellets that make up the rough interface because it is believed that these pellets may cause disturbances by generating noise. However, the results were not as expected. The classification with this new data only yielded similar results, with slight improvements or even sometimes less good results.

Table 4.7: Classification results with random forest for subtracted blanks (a) and normal data (b).

(a) Random forest, rough interface proud, blanks subtracted

	Predicted Elastic	Predicted Rigid
True Elastic	68	172
True Rigid	14	264

Metric	Value
Recall	0.28
Precision	0.83
Accuracy	0.66

(b) Random forest, rough interface proud, no subtraction of the blanks

	Predicted Elastic	Predicted Rigid
True Elastic	62	178
True Rigid	13	265

Metric	Value
Recall	0.26
Precision	0.83
Accuracy	0.65

Table 4.7 shows an example where the removal of the blanks showed an improvement, though very slight. The classifier placed six more targets in the elastic category than in the rigid category, while they were actually elastic. But it also wrongly placed one more rigid target in the wrong category than with normal data.

4.9 Logistic regression

The logistic regression algorithm, as explained in section 3.2.3.3, is a simpler algorithm than the other two used until now and should, therefore, yield weaker results than both of them. The algorithm was trained in the exact same way then for the other two, on the features extracted from the data in time domain and with filtered features. This was proven to be true for all environments, but with slight variations. Variations in the accuracy ranged from about one percent to five percent.

Table 4.8: Comparison of the results from logistic regression and XGBoost classifications, for free-field (a and b) and partially buried in rough interface (c and d).

(a) Logit regression, free-field

	Predicted Elastic	Predicted Rigid
True Elastic	509	18
True Rigid	14	294

Metric	Value
Recall	0.96
Precision	0.97
Accuracy	0.96

(b) XGBoost, free-field

	Predicted Elastic	Predicted Rigid
True Elastic	510	17
True Rigid	11	297

Metric	Value
Recall	0.97
Precision	0.98
Accuracy	0.97

(c) Logit regression, rough interface

	Predicted Elastic	Predicted Rigid
True Elastic	224	226
True Rigid	105	231

Metric	Value
Recall	0.50
Precision	0.68
Accuracy	0.58

(d) XGBoost, Rough interface

	Predicted Elastic	Predicted Rigid
True Elastic	215	235
True Rigid	72	264

Metric	Value
Recall	0.48
Precision	0.75
Accuracy	0.61

Table 4.8 shows some examples of logistic regression classification compared to the results of XGBoost for the same experimental conditions. Logistic regression consistently yielded lower performance than the other two classifiers.

4.10 Confidence intervals

Basic binary classification metrics such as recall, precision, the F1-score, or the ROC-AUC are just results for one sample of the global data population. They are already useful, but repetition on several other samples can expand the understanding of global data [92]. Variations in results for a given metric and a given classifier can be viewed with Confidence Interval (CI), that shows the scope of all the results and the mean value and other helpful statistics. To build a CI, classification is performed for several samples, and the results are gathered to find outliers and other metrics. In a way, this is a good tool to see if a classification algorithm is consistent and stable throughout the data and the measured metrics. Therefore, the narrower the CI's the better. The results for some CI are shown in figures 4.6 and 4.7, for the two most extreme cases (the best and worst possible environments tested, respectively). For targets placed on the flat interface (figure 4.6), The narrowest and highest CI is always the ROC-AUC metric for all three classifiers, while the widest is the recall. precision and the F1-score are equivalent in width and are intermediate between the F1-score and the ROC-AUC metrics. All CI metrics are also compromised in between 0.95 and 1.00. Regarding targets partially buried in the rough interface, the widths of all CIs appear more similar to one another, though the range spans from approximately 0.5 to 0.9

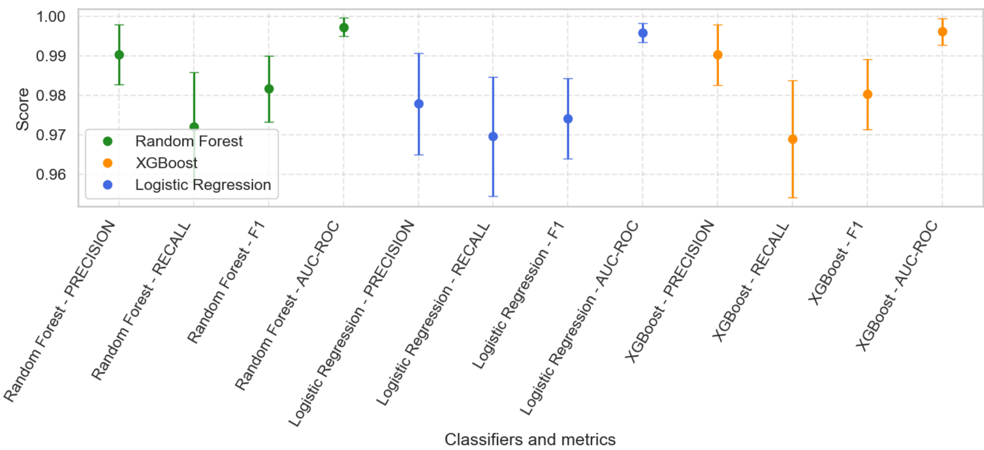

Figure 4.6: Confidence intervals for all three classifiers, for flat interface environment and filtered features.

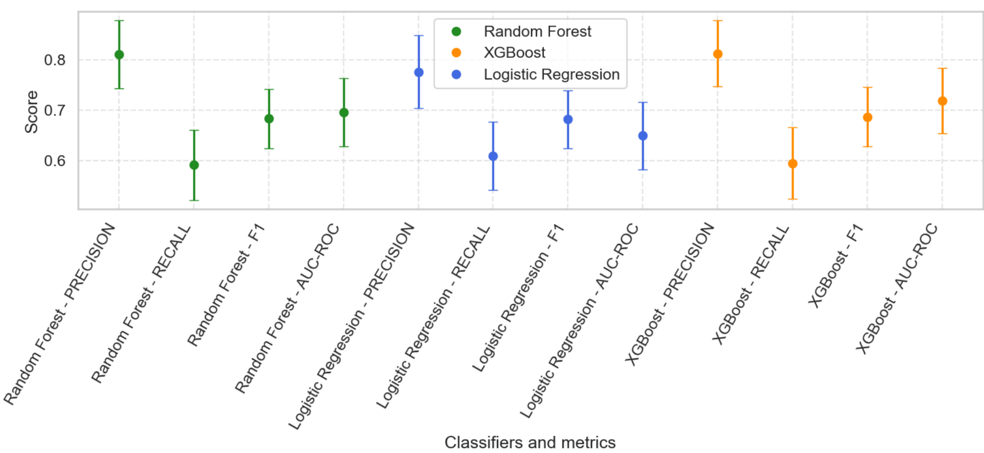

Figure 4.7: Confidence intervals for all three classifiers, for rough interface partially buried environment and filtered features.

Chapter 5

Discussion

Data classification can be performed in many different ways, and it is important to find the classification method that suits the experimental data the best. In this study, it was decided to use the time series to perform classification on targets detected by a sonar in air, which is a simplification of the underwater conditions encountered in real mine-countermeasure operations, by extracting features that explained the echoes as much as possible. Although this method seems promising for cases where the targets are not partially hidden or surrounded by other materials. The classifiers show some limitations when there is non negligible noise generated by the surroundings of the targets. It suggests that advanced data processing techniques are required in order to get rid of such noise or to find ways of extracting the useful data only and discarding the noise. An attempt was made to remove the noise by subtracting the blanks, but this proved unsuccessful because the performances were not enhanced as expected. The present results partially confirm the findings of Zerr et al. [45], who demonstrated the potential of resonance scattering for mine classification. However, our results indicate that these resonance signatures alone may not be sufficiently robust when environmental complexity increases, particularly under rough-interface conditions

It seems that the noise caused by the plastic pellets affects all parts of the time series in a way that the data becomes biased toward the rigid class, this is particularly visible in figure 4.2, that compares the recall, which is always high, but more importantly, the lower the recall of the elastic class the higher the recall of the rigid class. In fact, the plastic beads are very small, almost sand-like particles. The ka coefficient for particles of this size, the coefficient (Section 2.0.2) indicates that only specular reflections are expected upon insonification, which would be an explanation as to why the majority of the targets are classified as rigid for the complex environments. Another explanation for this bias could be the unbalance of the data. sometimes, there were far more data for the rigid class. The classifier then naturally tends to favor this specific class. To improve classification performance, the dataset should be balanced and the pellet-induced acoustic contributions suppressed.

Furthermore, The data under study was controlled and localized, with the position of the targets always known, facilitating the extraction of the windows in the processing of the data. In real conditions, the positions of the targets would never be known as precisely as in the controlled dataset used in this study. Other ways of data extraction would be required than simply calculating a time of flight with the known positions of the targets and the sonar. For instance, by reconstructing the high resolution SAS image, and localizing the possible targets with recognition algorithms, such as deep learning algorithms. But the same problem remains with this methodology. The current data in this field is too limited, and deep learning methods require large datasets on which to train. Otherwise, they are practically unusable. However, the possibility of using simulated data could be a solution to help generate more data for this purpose. Some studies demonstrated this possibility and the fact that training the algorithms with simulated data could be beneficial without necessarily degrading classification performance on real data. Another problem with using SAS, is that higher frequencies of sonar pulses are needed to get precise images, but to get the most out of the time series and the internal reflections of the targets that are partially buried or completely buried in the seabed for feature extraction and classification, lower frequencies are required. Lower frequencies are the ones that are able to penetrate through sand and generate echoes for buried targets. Another way of finding the targets in the whole collected data is to use algorithms that can detect spikes in the time series, corresponding to the reflections of the targets. The Constant Constant False rate Alarm (CFAR) is an algorithm that is able to do such tasks, by sequentially analyzing the echoes and searching for spikes of higher magnitude.

Another limitation that comes with the experimental setup is that because the experiments were performed in air rather than in water, the acoustic impedance contrast differs substantially from operational mine-hunting conditions. Consequently, the relative contribution of specular and resonance scattering mechanisms may differ from those expected in real underwater environments. A natural question when interpreting these results is whether a similar experiment conducted in water would yield comparable, better, or worse classification performance. From an acoustic physics standpoint, the key difference lies in impedance. The acoustic impedance of a medium is defined as Z=ρc, where ρ is the density and c is the speed of sound, expressed in Rayl (kg/mšůs). Water has an impedance of approximately 1.5 MRayl, while air is roughly 0.0004 MRayl a factor of around 3600 difference. In water, the impedance of materials such as aluminum or polyurethane is much closer to that of the surrounding medium, allowing the incident wave to partially penetrate the target and excite internal resonances that re-radiate as part of the echo. In air, the impedance contrast is so extreme that nearly all energy reflects specularly at the surface, with little coupling into the target interior. This suggests that elastic resonance signatures would likely be more pronounced in an underwater scenario, potentially making the rigid-vs-elastic classification task easier from a physical standpoint. However, this would need to be weighed against the increased environmental variability, noise present in real underwater deployments and complex internal structure and materials of mines and UXOs. So, even though the ka coefficient calculated for the data is at the right level for the appearance of internal reflections and vibrations, these signatures are not as present as in underwater environments. Data therefore have to be used with care.

One particular aspect that can be discussed is the fact that under water, the acoustic wave has to travel several meters downward and upward, possibly traversing multiple layers of waters with different densities. There are three main parameters that cause these density variations: water temperature, salinity, and pressure. This is important to take into account because the speed of sound in a medium depends on its composition, but also on its density for a fluid such as water, and if there is a variation of speed, the sound wave will bend. The question then is: does this impact UXO survey in sea waters? The data used in this study were collected with the sonar being close to the targets. And in the air, the variation of the speed of sound is not as important as in a much denser fluid that is water. So, this variation of density and the aspect of a layered medium is not encapsulated in the data used here. In real applications, the distance separating a sonar from its targets varies a lot. Low frequency sonars mounted on the hull of a boat can reach far greater distances, so the sound speed profile of the water should therefore be taken into account. But for AUV's, that can navigate more easily through the water are able to get very close to their targets. The sound speed profile can therefore be ignored. So this parameter really depends on the specific application being considered and the type of data required.

Filtering of the features was proven to be adequate. The classification power was almost not altered by reducing the features from more than 2300 to 198, allowing for faster calculations and less resources needed. Furthermore, the extraction of features on all 3 components of the complex signal (real and imaginary parts and the envelope) allowed to triple the number of available useful features.

Moreover, the experiments were conducted with only four target types, only two of which - the two spheres - that resemble some real underwater mines and **??**s. A more rigorous study should have additional target shapes, materials and dimensions, that look like real mines found at sea to assess the robustness of the proposed methodology.

Comparing the three classifiers showed that logistic regression, which has a simpler architecture, consistently underperformed the two decision-tree-based methods. For further research, this type of algorithm could be left behind.

Chapter 6

Conclusion and prospects

6.1 Conclusion

There is an urgent need to accelerate the neutralization of underwater mines and UXOs, as they remain a threat to marine and human life, as their explosive charges may still be active even when they have been present in the waters for a long time. With humans building more and more offshore structures, such as oil rigs or offshore wind farms, there is also a demand to protect the areas under construction. This is why there exist laws, jurisdictions, and framework directives pressing countries to act fast in the dismantling of bombs. Some efforts have already been made and new technologies arise to reach and neutralize UXO's faster and cheaper than before, like with the use of AUV's carrying sonar systems.

In this process of UXO management, one crucial part is to classify the data that were collected with sonar technologies. The need to be able to discriminate a mine from every other object that is not a threat to the environment and fauna is really important because surveys cost a lot of money and resources. Accurate classification translates into significant gains in operational efficiency and it is important to be able to correctly classify a rock as a rock and not as a bomb, because it would be a waste of resources to reach for a rock and notice that it was actually not a threat. Furthermore, another key aspect of classification is to avoid the false classification of a mine as another object, leaving a potential dangerous ordnance in the environment for good. This idea is at the core of classification optimization: avoiding false positives and false negatives.

Classification was performed on data collected by [15], a research team from the University of New Hampshire in the United States. Their original goal was to collect acoustic data to perform SAS processing to build high resolution images of the insonified targets on several environments, including free-field, proud on a flat interface, proud or partially buried in a rough interface made of plastic pellets. Four targets were studied: a polyurethane sphere, an aluminum sphere, and two MDF letters, an O and a Q. these targets can be placed into two categories: the spheres act as elastic reflectors, and the letters act as rigid reflectors toward acoustic waves. The data was used just before the SAS processing, but after a series of data processing involving several filters, a Hilbert transform, and a matched filter to get a complex signal with enhanced acoustic signatures. After data processing, the echoes of the targets had to be cut from the overall data, and from these windows of time series, or echoes, features were extracted with tsfresh. These features were filtered by a process of correlation followed by a cross-selection, to keep the most relevant ones, for performance enhancement and computing efficiency. These filtered features were then fed to supervised machine learning algorithms to perform classification. Three different algorithms were used. On the one hand, two algorithms based on decision trees and, on the other hand, a logistic regression algorithm. Classification was performed on two classes of targets: elastic targets mimicking man made objects, such as metallic underwater bombs and mines, and rigid targets mimicking rocks. The difference in behavior between these two classes of objects is that an acoustic wave hitting an elastic object will generate internal vibrations within the target, which will then be scattered back to the surrounding fluid (in the air in this case). Whereas, when a rigid object is hit by an acoustic wave, the incoming beam will only reflect off of the surface, without internal vibrations. This is exactly what the classifiers were meant to capture, using features to describe the echoes.

Blanford et al. [15] (2024) used the dataset of raw acoustic signals, in the form of time series, to recreate high resolution images of different targets. But visual interpretation on such images tends to be limited to distinguish targets that have similar shapes. This is why, in this paper, the potential of the raw acoustic data was used to gather more information on the internal structure of the targets, thanks to their ability to vibrate when they are insonified by an acoustic wave sent by a sonar. This structural behavior is at the core of the classification methodology, as some targets behave more like elastic objects and therefore emit more vibrations to the environment. This is a concept that is completely ignored in the study of Blanford et al. [15].

The results varied a lot from environment to environment. Classification performed for easier environments (free-field and flat interface) yielded very good results, with accuracy levels reaching almost 100% in some cases. However, for the complex environment (rough interface), the results were poorer, with a tendency to classify most of the targets as rigid. To counteract this problem, an attempt was made to subtract the blanks (scans of the scene without the targets) from the actual scans, in order to only keep the contribution of the targets to the echoes, by removing potential noise generated by the plastic pellets making up the rough interface. This was an unsuccessful attempt, as the classification results were marginally changed after blank removal, sometimes a little better, sometimes a little worse than before blank removal.

The comparison of the results from the different classifiers showed that both algorithms relying on decision trees, the random forest and XGBoost, had similar results in most cases, with XGBoost having the better overall results by a small margin. The logistic algorithm was constantly behind the other two.

This study explored classification further than previous studies on the same topic, as depicted in section 2, in the sense that previous studies only performed classification on either one target type, or one environment type. The strength of this paper lies in the unprecedented combination of target types and environmental variability. Moreover, the use of a fast features extractor, TSfresh, for this purpose was not done yet. Most studies only used a limited amount of features, while in this study, several dozens were calculated, giving more information on the signals for the classificators.

6.2 Prospects

The classification results obtained in this study lead to the conclusion that using the time series might not be the go-to solution to classify underwater targets. However, these results need to be put into perspective. Data were taken in air and not in water; therefore, the data is less exploitable. Some of the results were also very promising (targets in free-field and flat interface conditions). This shows the potential of machine learning algorithms in the field, but the actually most of the UXOs and naval mines in nature are either lying proud on the surface of the seabed, in some cases even covered in vegetation and other debris, or partially or completely buried in the seabed. Conditions for which the classification results were far too low to be exploitable. This is why feature-based machine learning applied to time-series data is not recommended for complex environments when measurements are performed in air.

For further studies, data collected in the water are needed as the elastic behavior might be more significant in the time series, with more pronounced spikes and scattering echoes, thus not being hindered by the specular reflections caused by a rough interface and also being more comparable with real applications.

A combination of classification techniques could also yield more suitable results, combining SAS image-based visual detection potential UXOs with feature-based deep learning classification methods, but this would require a larger amount of data than is currently available online. To address this problem, data simulation to enhance the quantity of overall data seems to be a good lead.

Bibliography

[1] Edmund Maser and Jennifer S. Strehse. dont blast: blast-in-place (BiP) operations of dumped world war munitions in the oceans significantly increase hazards to the environment and the human seafood consumer. 94(6):1941–1953. ISSN 1432-0738. doi: 10.1007/s00204-020-02743-0. URL https://doi.org/10.1007/s00204-020-02743-0.

[2] The allied north sea mine barrage of world war i, . URL https://www.abmc.gov/news-events/news/allied-north-sea-mine-barrage-world-war-i/.

[3] James Simpson. How britain beat germanys magnetic sea mines https://medium.com/war-is-boring/how-britain-beat-germanys-wwii-magnetic-sea-mines-bfec5558704c. URL https://medium.com/war-is-boring/how-britain-beat-germanys-wwii-magnetic-sea-mines-bfec5558704c.

[4] Mine warfare at sea in a korea contingency korea institute for maritime strategy, . URL https://en.kims.or.kr/issubrief/kims-periscope/peri301/.

[5] Mines of russia / USSR - NavWeaps http://www.navweaps.com/weapons/WAMRussian_mines.php, . URL http://www.navweaps.com/Weapons/WAMRussian_Mines.php.

[6] Nick Cooper and Simon Cooke. Risky business: Dealing with unexploded ordnance (UXO) in the marine environment. In Kevin Burgess, editor, *Coasts, Marine Structures and Breakwaters 2017*, page 0. ICE Publishing. ISBN 978-0-7277-6317-4. doi: 10.1680/cmsb.63174.0157. URL https://doi.org/10.1680/cmsb.63174.0157.

[7] Joseph Needham. Science and civilization in china. volume 5. chemistry and chemical technology. part 7: Military technology; the gunpowder epic. doi: 10.1017/S0035869X0016486X. URL https://www.researchgate.net/publication/273538092_Science_and_civilization_in_China_Volume_5_Chemistry_and_chemical_technology_Part_7_Military_technology_the_gunpowder_epic_By_Joseph_Needham_with_the_collaboration_of_Ho_Ping-Yu_Lu_Gwei-Djen_and_Wang_.

[8] Robert Routledge. *Discoveries and Inventions of the Nineteenth Century*. URL https://www.gutenberg.org/ebooks/54475.

[9] Jean-Paul Ducrotoy, Mike Elliott, and Victor N de Jonge. The north sea. 41(1):5–23. ISSN 0025-326X. doi: 10.1016/S0025-326X(00)00099-0. URL https://www.sciencedirect.com/science/article/pii/S0025326X00000990.

[10] Peter Menzel, Anja Drews, Tjark Mehring, Christoph Otto, and Dorthe Reng Erbs-Hansen. Mobilization of unexploded ordnance on the seabed. 10(7):389. ISSN 2305-6304. doi: 10.3390/toxics10070389. URL https://www.mdpi.com/2305-6304/10/7/389.

[11] Spencer Kimball. Iran has mined 'large segments' of hormuz strait, secretary of state rubio says. URL https://www.cnbc.com/2026/06/02/iran-war-strait-hormuz-marco-rubio-mine.html. Section: Energy.

[12] Explosion kills three and sinks ukrainian dredge on a danube canal, . URL https://maritime-executive.com/article/explosion-kills-three-and-sinks-ukrainian-dredge-on-a-danube-canal.

[13] Par Victor Cousin Le 13 mars 2026 à 06h00. Guerre en iran : ń lune des armes les plus destructrices jamais vues ż, cest quoi les mines maham déployées dans le détroit dormuz ? URL https://www.leparisien.fr/international/iran/guerre-en-iran-lune-des-armes-les-plus-destructrices-jamais-vues-cest-quoi-les-mines-maham-deployees-dans-le-detroit-dormuz-13-03-2026-RBT4K5XY4VAX5I6IJKZ7KSIC7E.php. Section: /international/iran/.

[14] Christopher McFadden. Why submerged munitions still matter decades after war. URL https://interestingengineering.com/military/submerged-munitions-contamination-marine-ecosystem.

[15] Thomas E. Blanford, David P. Williams, J. Daniel Park, Brian T. Reinhardt, Kyle S. Dalton, Shawn F. Johnson, and Daniel C. Brown. An in-air synthetic aperture sonar dataset of target scattering in environments of varying complexity. 11(1):1196. ISSN 2052-4463. doi: 10.1038/s41597-024-04050-0. URL https://www.nature.com/articles/s41597-024-04050-0.

[16] radar used underwater instead of sonar if radio waves are faster than sound waves?, . URL https://www.physlink.com/education/askexperts/ae456.cfm.

[17] Relationship between antenna size and wavelength_industry news_news_shenzhen VLG WIRELESS TECHNOLOGY CO., LTD, . URL https://www.vlg-solution.com/industry_news/Relationship_between_Antenna_Size_and_Wavelength_497.html.

[18] Keir Colbo, Tetjana Ross, Craig Brown, and Tom Weber. A review of oceanographic applications of water column data from multibeam echosounders. 145:41–56. ISSN 0272-7714. doi: 10.1016/j.ecss.2014.04.002. URL https://www.sciencedirect.com/science/article/pii/S0272771414000900.

[19] Chirp sub-bottom profiler processing-a review - ProQuest, . URL https://www.proquest.com/docview/198554237?sourcetype=Trade%20Journals.

[20] Philippe Blondel. The handbook of sidescan sonar. pages 249–276. ISBN 978-3-540-42641-7. doi: 10.1007/978-3-540-49886-5_11.

[21] Peter Gough. SYNTHETIC APERTURE SONAR: THE PAST, THE PRESENT AND THE FUTURE. In *ResearchGate*. doi: 10.25144/18014. URL https://www.researchgate.net/publication/379532257_SYNTHETIC_APERTURE_SONAR_THE_PAST_THE_PRESENT_AND_THE_FUTURE.

[22] D. Marx, M. Nelson, E. Chang, W. Gillespie, A. Putney, and K. Warman. An introduction to synthetic aperture sonar. In *Proceedings of the Tenth IEEE Workshop on Statistical Signal and Array Processing (Cat. No.00TH8496)*, pages 717–721. doi: 10.1109/SSAP.2000.870220. URL https://ieeexplore.ieee.org/abstract/document/870220.

[23] Trevor Hastie, Robert Tibshirani, Jerome Friedman, and James Franklin. The elements of statistical learning: Data mining, inference, and prediction. 27:83–85. doi: 10.1007/BF02985802.

[24] R. Saravanan and Pothula Sujatha. A state of art techniques on machine learning algorithms: A perspective of supervised learning approaches in data classification. In *2018 Second International Conference on Intelligent Computing and Control Systems (ICICCS)*, pages 945–949. doi:

10.1109/ICCONS.2018.8663155. URL https://ieeexplore.ieee.org/abstract/document/8663155.

[25] Reza Ravanmehr and Rezvan Mohamadrezaei. Deep learning overview. In Reza Ravanmehr and Rezvan Mohamadrezaei, editors, *Session-Based Recommender Systems Using Deep Learning*, pages 27–72. Springer Nature Switzerland. ISBN 978-3-031-42559-2. doi: 10.1007/978-3-031-42559-2_2. URL https://doi.org/10.1007/978-3-031-42559-2_2.

[26] Peter McCullagh and Jie Yang. Stochastic classication models. In *Proceedings of the International Congress of Mathematicians Madrid, August 2230, 2006*, pages 669–686. European Mathematical Society - EMS - Publishing House GmbH. doi: 10.4171/022-3. URL https://ems.press/books/standalone/24/527.

[27] Biao Qin, Yuni Xia, Sunil Prabhakar, and Yicheng Tu. A rule-based classification algorithm for uncertain data. In *2009 IEEE 25th International Conference on Data Engineering*, pages 1633–1640. doi: 10.1109/ICDE.2009.164. URL https://ieeexplore.ieee.org/abstract/document/4812586. ISSN: 2375-026X.

[28] Herman Medwin and Clarence S. Clay. Chapter 7 - sound scattered by a body. In *Fundamentals of Acoustical Oceanography*, Applications of Modern Acoustics, pages 234–286. Academic Press, . ISBN 978-0-12-487570-8. doi: 10.1016/B978-012487570-8/50009-X. URL https://www.sciencedirect.com/science/article/pii/B978012487570850009X.

[29] Herman Medwin and Clarence Samuel Clay. *Fundamentals of acoustical oceanography*. Applications of modern acoustics. Academic Press, . ISBN 978-0-12-487570-8.

[30] SPF Santé publique. Convention des nations unies sur le droit de la mer (UNCLOS) | SPF santé publique. URL https://www.health.belgium.be/fr/organisation-politique/reglementation-documents-politique/convention-nations-unies-droit-mer-unclos.

[31] Protocol on explosive remnants of war (protocol v to the 1980 CCW convention), 28 november 2003, . URL https://ihl-databases.icrc.org/en/ihl-treaties/ccw-protocol-v.

[32] EU marine strategy framework directive - research and innovation, . URL https://research-and-innovation.ec.europa.eu/research-area/environment/oceans-and-seas/eu-marine-strategy-framework-directive_en.

[33] Système de lutte anti-mines du futur (slamf)ă: positionner la france parmiălesăleaders desădronesănavals, . URL https://www.defnat.com/e-RDN/vue-article-cahier.php?carticle=54&cidcahier=1161.

[34] Inertial navigation systems (INS) an introduction, . URL https://www.advancednavigation.com/tech-articles/inertial-navigation-systems-ins-an-introduction/.

[35] Qinghua Luo, Xiaozhen Yan, Chunyu Ju, Yunsai Chen, and Zhenhua Luo. An ultra-short baseline underwater positioning system with kalman filtering. 21(1):143. ISSN 1424-8220. doi: 10.3390/s21010143. URL https://pmc.ncbi.nlm.nih.gov/articles/PMC7796008/.

[36] National Research Council (U S.) Committee for Mine Warfare Assessment and National Research Council (U.S.). Committee for Mine Warfare Assessment. *Naval mine warfare : operational and technical challenges for naval forces*. National academy press. ISBN 978-0-309-07578-7.

[37] le marin. La marine nationale soigne son poisson autopropulsé. URL https://lemarin.ouest-france.fr/defense/la-marine-nationale-soigne-son-poisson-autopropulse-37ddee27-8977-4a47-87c8-a1d60ca22117. Section: Défense.

[38] E. M. Lifshitz, A. M. Kosevich, and L. P. Pitaevskii. CHAPTER i - FUNDAMENTAL EQUATIONS. In *Theory of Elasticity (Third Edition)*, pages 1–37. Butterworth-Heinemann. ISBN 978-0-08-057069-3. doi: 10.1016/B978-0-08-057069-3.50008-5. URL https://www.sciencedirect.com/science/article/pii/B9780080570693500085.

[39] Mike Ashby, Hugh Shercliff, and David Cebon. Chapter 10 - fracture- and fatigue-limited design. In *Materials (Fourth Edition)*, pages 255–280. Butterworth-Heinemann. ISBN 978-0-08-102376-1. doi: 10.1016/B978-0-08-102376-1.00010-X. URL https://www.sciencedirect.com/science/article/pii/B978008102376100010X.

[40] B.H. Houston, J.A. Bucaro, T. Yoder, L. Kraus, J. Tressler, J. Fernandez, T. Montgomery, and T. Howarth. Broadband low frequency sonar for non-imaging based identification. In *OCEANS '02 MTS/IEEE*, volume 1, pages 383–387 vol.1. doi: 10.1109/OCEANS.2002.1193302. URL https://ieeexplore.ieee.org/abstract/document/1193302.

[41] Mariia Dmitrieva, Keith Brown, Gary Heald, and David Lane. Material recognition based on the time delay of secondary reflections using wideband sonar pulses. 13(11):2034–2040. ISSN 1751-8792. doi: 10.1049/iet-rsn.2018.5493. URL https://onlinelibrary.wiley.com/doi/abs/10.1049/iet-rsn.2018.5493. _eprint: https://ietresearch.onlinelibrary.wiley.com/doi/pdf/10.1049/iet-rsn.2018.5493.

[42] Jianer Huang and Xiukun Li. Radius estimation of thin spherical shells proud on the sediment based on resonance peak intervals. In *2021 OES China Ocean Acoustics (COA)*, pages 127–130. doi: 10.1109/COA50123.2021.9519888. URL https://ieeexplore.ieee.org/document/9519888.

[43] Tianyang Xu, Xiukun Li, and Hongjian Jia. Underwater spherical shell classification and parameter estimation based on acoustic backscattering characteristics. 9:162756–162764. ISSN 2169-3536. doi: 10.1109/ACCESS.2020.3046364. URL https://ieeexplore.ieee.org/document/9301280.

[44] Active sonar target classification with power-normalized cepstral coefficients and convolutional neural network, . URL https://www.mdpi.com/2076-3417/10/23/8450.

[45] B. Zerr, A. Tesei, A. Maguer, W. L. Fox, and J. A. Fawcett. A classification technique combining aspect dependence and elastic properties of target scattering. URL https://apps.dtic.mil/sti/html/trecms/AD1113989/.

[46] Hongjian Jia, Xiukun Li, and Xiangxia Meng. Rigid and elastic acoustic scattering signal separation for underwater target. 142(2):653–665. ISSN 0001-4966, 1520-8524. doi: 10.1121/1.4996127. URL https://pubs.aip.org/jasa/article/142/2/653/640423/Rigid-and-elastic-acoustic-scattering-signal.

[47] N. Dasgupta, P. Runkle, L. Carin, L. Couchman, T. Yoder, J. Bucaro, and G.J. Dobeck. Class-based target identification with multiaspect scattering data. 28(2):271–282. ISSN 1558-1691. doi: 10.1109/JOE.2003.811899. URL https://ieeexplore.ieee.org/document/1209626.

[48] Yoojeong Seo, Baeksan On, Sungbin Im, Taebo Shim, and Iksu Seo. Underwater cylindrical object detection using the spectral features of active sonar signals with logistic regression models. 8(1):116. ISSN 2076-3417. doi: 10.3390/app8010116. URL https://www.mdpi.com/2076-3417/8/1/116.

[49] W. Hume-Rothery. Elasticity and anelasticity of metals. 164(4159):84–85. ISSN 1476-4687. doi: 10.1038/164084b0. URL `https://www.nature.com/articles/164084b0`.

[50] C. P. Buckley, C. Prisacariu, and C. Martin. Elasticity and inelasticity of thermoplastic polyurethane elastomers: Sensitivity to chemical and physical structure. 51(14):3213–3224. ISSN 0032-3861. doi: 10.1016/j.polymer.2010.04.069. URL `https://www.sciencedirect.com/science/article/pii/S0032386110003964`.

[51] Michael Rubinstein and Sergei Panyukov. Elasticity of polymer networks. 35(17):6670–6686. ISSN 0024-9297. doi: 10.1021/ma0203849. URL `https://doi.org/10.1021/ma0203849`.

[52] Feng Kang and Shi Zhong-Ci. *Mathematical Theory of Elastic Structures*. Springer. ISBN 978-3-662-03288-6 978-3-662-03286-2. doi: 10.1007/978-3-662-03286-2. URL `http://link.springer.com/10.1007/978-3-662-03286-2`.

[53] Elastic deformation: Hooke's law and stiffness, . URL `https://www.doitpoms.ac.uk/tlplib/mechanical_properties/elastic.php`.

[54] K. E. Gilbert and H. E. Bass. ACOUSTIC WAVES. In James R. Holton, editor, *Encyclopedia of Atmospheric Sciences*, pages 1–12. Academic Press. ISBN 978-0-12-227090-1. doi: 10.1016/B0-12-227090-8/00046-4. URL `https://www.sciencedirect.com/science/article/pii/B0122270908000464`.

[55] Jianfeng Zhang. Wave propagation across fluid-solid interfaces: A grid method approach. 159: 240–252. doi: 10.1111/j.1365-246X.2004.02372.x.

[56] Chapter 6 reflection and transmission of plane waves. In *Handbook of Geophysical Exploration: Seismic Exploration*, volume 31, pages 167–218. Pergamon, . doi: 10.1016/S0950-1401(01)80044-6. URL `https://www.sciencedirect.com/science/chapter/handbook/abs/pii/S0950140101800446`.

[57] Houjiang Wang, Wei Lin, and Jincheng Gu. Study on one-way transmission of acoustic wave based on metasurface. 1978:012025, . doi: 10.1088/1742-6596/1978/1/012025.

[58] B. D. Zaitsev, I. A. Borodina, A. A. Teplykh, and A. P. Semyonov. Determination of the acoustic wave velocity and attenuation in liquids with different acoustic impedances using an acoustic interferometer. 69(4):503–509. ISSN 1562-6865. doi: 10.1134/S1063771023600493. URL `https://doi.org/10.1134/S1063771023600493`.

[59] Xiaochan Wang, Jianxin Ma, Qi Zhang, and Xiangjun Xin. Generation of linear frequency-modulated signals with improved time-bandwidth product based on an optical frequency comb. 58:3222–3228, . doi: 10.1364/AO.58.003222.

[60] Nadav Levanon and Eli Mozeson. *Radar Signals*. ISBN 978-0-471-47378-7. doi: 10.1002/0471663085.

[61] Jennifer E. Michaels, Sang Jun Lee, Anthony J. Croxford, and Paul D. Wilcox. Chirp excitation of ultrasonic guided waves. 53(1):265–270. ISSN 0041-624X. doi: 10.1016/j.ultras.2012.06.010. URL `https://www.sciencedirect.com/science/article/pii/S0041624X12001333`.

[62] Douglas Lyon. The discrete fourier transform, part 4: Spectral leakage. 8(7). doi: 10.5381/jot.2009.8.7.c2. URL `https://digitalcommons.fairfield.edu/engineering-facultypubs/73`.

[63] K.M.M. Prabhu. *Window Functions and Their Applications in Signal Processing*. ISBN 978-1-4665-1583-3. doi: 10.1201/9781315216386.

[64] David P. Williams. Transfer learning with SAS-image convolutional neural networks for improved underwater target classification. In *IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium*, pages 78–81, . doi: 10.1109/IGARSS.2019.8898611. URL https://ieeexplore.ieee.org/abstract/document/8898611. ISSN: 2153-7003.

[65] David P. Williams. Underwater target classification in synthetic aperture sonar imagery using deep convolutional neural networks. In *2016 23rd International Conference on Pattern Recognition (ICPR)*, pages 2497–2502, . doi: 10.1109/ICPR.2016.7900011. URL https://ieeexplore.ieee.org/abstract/document/7900011.

[66] Avi Abu and Roee Diamant. Underwater object classification combining SAS and transferred optical-to-SAS imagery. 144:109868. ISSN 0031-3203. doi: 10.1016/j.patcog.2023.109868. URL https://www.sciencedirect.com/science/article/pii/S0031320323005666.

[67] Stephen Elliott, Mihai Orita, Erika Quaranta, and Jordan Cheer. Acoustic scattering from spherical shells and its active control. 568:118056. ISSN 0022-460X. doi: 10.1016/j.jsv.2023.118056. URL https://www.sciencedirect.com/science/article/pii/S0022460X23005059.

[68] Kevin Krout and Siavash Sohrab. On the speed of sound. 19:29. doi: 10.5541/ijot.5000155148.

[69] Robert Balmer. *Modern Engineering Thermodynamics*. Academic Press. ISBN 978-0-08-096173-6.

[70] Owen Cramer. The variation of the specific heat ratio and the speed of sound in air with temperature, pressure, humidity, and CO2 concentration. 93(5):2510–2516. ISSN 0001-4966. doi: 10.1121/1.405827. URL https://doi.org/10.1121/1.405827.

[71] tblanford. tblanford/airsas. URL https://github.com/tblanford/airsas. original-date: 2024-03-13T17:07:44Z.

[72] Nader Jalali Marand, Hassan Sam Daliri, and Hossein Askarian-Abyaneh. Removal of DC offset components by using two auxiliary signals. In *2023 IEEE PES GTD International Conference and Exposition (GTD)*, pages 226–230. doi: 10.1109/GTD49768.2023.00069. URL https://ieeexplore.ieee.org/document/10261736.

[73] E. J. Hannan and P. J. Thomson. Estimating group delay. 60(2):241–253. ISSN 0006-3444, 1464-3510. doi: 10.1093/biomet/60.2.241. URL https://academic.oup.com/biomet/article-lookup/doi/10.1093/biomet/60.2.241.

[74] Carlos J. Chavez, Kent D. Benson, Robert K. Fahler, Robert J. Frank, and Scott J. F. Zogg. Signal acquisition with transmit blanking compensation. URL https://patents.google.com/patent/US7508887B1/en.

[75] D. Ibrahim. Accurate measurement of the mains electricity frequency. URL https://www.semanticscholar.org/paper/Accurate-Measurement-of-the-Mains-Electricity-Ibrahim/0dd8de1a36e814f12fcf0fe6612df44772ea3154.

[76] Mains frequency, . URL https://www.power2market.com/en/news/netzfrequenz.

[77] Frederick W. King. *Hilbert Transforms*, volume 1 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press. ISBN 978-0-521-88762-5. doi: 10.1017/CBO9780511721458. URL https://www.cambridge.org/core/books/hilbert-transforms/9FEE707FB7587A9A1A90C0038FB77E09.

[78] J. Dugundji. Envelopes and pre-envelopes of real waveforms. 4(1):53–57. ISSN 2168-2712. doi: 10.1109/TIT.1958.1057435. URL https://ieeexplore.ieee.org/abstract/document/1057435.

[79] H. Poor. Robust matched filters. 29(5):677–687. ISSN 1557-9654. doi: 10.1109/TIT.1983.1056734. URL https://ieeexplore.ieee.org/abstract/document/1056734.

[80] Laurent Younes. Introduction to machine learning. URL http://arxiv.org/abs/2409.02668.

[81] Xue Ying. An overview of overfitting and its solutions. 1168(2):022022. ISSN 1742-6596. doi: 10.1088/1742-6596/1168/2/022022. URL https://doi.org/10.1088/1742-6596/1168/2/022022.

[82] TseKiChun. English: Random forest explain. URL https://commons.wikimedia.org/wiki/File:Random_forest_explain.png.

[83] Erwan Scornet and Giles Hooker. Theory of random forests. 13:99–121. ISSN 2326-831X. doi: 10.1146/annurev-statistics-112723-034707. URL https://www.annualreviews.org/content/journals/10.1146/annurev-statistics-112723-034707.

[84] Tongtian Zhu. Analysis on the applicability of the random forest. 1607(1):012123. ISSN 1742-6596. doi: 10.1088/1742-6596/1607/1/012123. URL https://doi.org/10.1088/1742-6596/1607/1/012123.

[85] Gérard Biau and Erwan Scornet. A random forest guided tour. 25(2):197–227. ISSN 1863-8260. doi: 10.1007/s11749-016-0481-7. URL https://doi.org/10.1007/s11749-016-0481-7.

[86] Tianqi Chen and Carlos Guestrin. *XGBoost: A Scalable Tree Boosting System.* doi: 10.1145/2939672.2939785. Pages: 794.

[87] Miao Zou, Wu-Gui Jiang, Qing-Hua Qin, Yu-Cheng Liu, and Mao-Lin Li. Optimized XGBoost model with small dataset for predicting relative density of ti-6al-4v parts manufactured by selective laser melting. 15(15):5298, . ISSN 1996-1944. doi: 10.3390/ma15155298. URL https://pmc.ncbi.nlm.nih.gov/articles/PMC9369844/.

[88] Why use XGBoost? speed and performance advantages, . URL https://apxml.com/courses/getting-started-with-gradient-boosting-algorithms/chapter-4-advanced-gradient-boosting-xgboost/why-xgboost.

[89] Xiaonan Zou, Yong Hu, Zhewen Tian, and Kaiyuan Shen. Logistic regression model optimization and case analysis. In *2019 IEEE 7th International Conference on Computer Science and Network Technology (ICCSNT)*, pages 135–139, . doi: 10.1109/ICCSNT47585.2019.8962457. URL https://ieeexplore.ieee.org/abstract/document/8962457.

[90] Senthil kumar Anantharaman. A literature review of logistic regression models and its limitations. URL https://papers.ssrn.com/abstract=5285624.

[91] Sathyanarayanan Swaminathan and B Roopashri Tantri. Confusion matrix-based performance evaluation metrics. 27:4023–4031. doi: 10.53555/AJBR.v27i4S.4345.

[92] Sean Wallis. Binomial confidence intervals and contingency tests: Mathematical fundamentals and the evaluation of alternative methods. 20(3):178–208. ISSN 0929-6174. doi: 10.1080/09296174.2013.799918. URL https://doi.org/10.1080/09296174.2013.799918. _eprint: https://doi.org/10.1080/09296174.2013.799918.

Machine learning classification of rigid and elastic sonar targets using backscattering features

Application to underwater mines detection and environmental risk mitigation in marine ecosystems

Reliable discrimination between rigid underwater targets, such as natural rocks, and elastic targets, including unexploded ordnance (UXO) and sea mines, is a fundamental challenge in active sonar-based survey operations. Elastic objects, such as the metallic shells typical of UXOs and mines, generate resonant acoustic echoes when insonified at acoustic frequencies whose wavelengths are comparable to the target dimensions (10–30 kHz in this study), producing backscatter signatures that differ from those of rigid scatterers. This work exploits these differences through a feature-based machine learning approach applied to a publicly available dataset of controlled in-air sonar measurements. The analysis is performed on backscattered acoustic time series after standard signal processing steps including matched filtering. Data from four targets — a solid polyurethane sphere and a hollow aluminum shell (elastic), and two medium-density fiberboard block letters (rigid) — acquired across four environments of increasing complexity (free-field, flat interface, rough interface with proud targets, and rough interface with partially buried targets) were used. Features were extracted from the real, imaginary, and envelope components of each echo window using the Python library tsfresh, and subsequently reduced via statistical feature selection. Classification was performed using Random Forest and XGBoost ensembles. In free-field and flat-interface conditions, both classifiers exceeded 0.97 accuracy. Performance degraded substantially in rough-interface scenarios, approximately 0.61 to 0.65, where surface clutter obscures the resonance signatures critical to classification. These results highlight the sensitivity of time-domain feature extraction to environmental conditions, motivating the development of more robust preprocessing strategies.

Nathan Struyf

UNIVERSITÉ CATHOLIQUE DE LOUVAIN
Faculté des bioingénieurs
Croix du Sud, 2 bte L7.05.01, 1348 Louvain-la-Neuve, Belgique | **www.uclouvain.be/agro**