Università degli Studi di Napoli "Federico II"

SCUOLA POLITECNICA E DELLE SCIENZE DI BASE

DIPARTIMENTO DI INGEGNERIA INDUSTRIALE

TESI DI LAUREA IN INGEGNERIA AEROSPAZIALE

CLASSE DELLE LAUREE MAGISTRALI IN INGEGNERIA AEROSPAZIALE E ASTRONAUTICA (LM 20)

Impact Detection in Complex Composite Structures using Acoustic Emission and Improved Sensing Logic

Relatore:
Prof. Eng. Ernesto Monaco

Candidato:
Luigi Russo

Correlatore/i:
Prof. Eng. Vittorio Memmolo
M.Sc. Ahmed Bayoumi
M.Sc. Maria Moix-Bonet
Prof. Dr.-Eng. Peter Wierach

ANNO ACCADEMICO 2025 – 2026

Abstract

In the aerospace industry, Structural Health Monitoring (SHM) systems are essential to optimize maintenance and prevent catastrophic failures. This work aims to improve the accuracy of damage localization using the Acoustic Emissions (AE) technique, overcoming the intrinsic limitations of the ΔT method applied to composite materials (CFRP) compared to isotropic materials. The main objective is to localize the source of the damage within a target area of approximately $20 \times 20mm$, developing and comparing the ΔT method with a Machine Learning algorithm based on Random Forest.
The experimental campaign, conducted at the DLR, analyzed two test plates, one made of aluminum and one made of composite material. On aluminum, the ΔT method confirmed high accuracy with an error of 5.61 mm. On the CFRP plate, the intended target was successfully achieved in the internal area circumscribed by the sensors, where the ΔT method recorded an average error of $17.92mm$ and a localization radius of 20.83 mm (with 95% confidence). Evaluating the entire surface of the CFRP plate, the Random Forest algorithm achieved an accuracy of 34.17 mm and 68.19 mm of radius, outperforming the standard ΔT on the global area. The results validate the effectiveness of the methodology in the area circumscribed by the sensor network, providing a solid basis for future implementations on more complex aerospace structures.

Nomenclature

Acronyms

SHM	Structural Health Monitoring
CFRP	Carbon Fiber Reinforced Polymer
MRO	Maintenance, Repair and Overhaul
BVID	Barely Visible Impact Damage
NDT	Non-Destructive Testing
AE	Acoustic emission
MARSE	Measured Area of Rectified Signal Envelope
RMS	Root Mean Square
DTM	Delta T Mapping
TOA	Time of Arrival
ASTM	American Society for Testing and Materials
AI	Artificial Intelligence
AR	Auto-Regressive
AIC	Akaike Information Criterion
MAIC	Maeda's Akaike Information Criterion
CWT	Continuous Wavelet Transform
QI	Quasi-Isotropic
POL	Probability of Localization
H-N	Hsu-Nielsen
ANN	Artificial Neural Network
DSP	Digital Signal Processing
MAF	Moving Average Filter
DWT	Discrete Wavelet Transform
FIR	Finite Impulse Response
SNR	Signal to Noise Ratio
RSS	Residual Sum of Squares
CWT	Continuous Wavelet Transform
CB	Confidence Bound

Symbols

A	Signal amplitude
N	Signal counts
Et	Absolute energy of a AE burst
R	Signal rise time
ΔT	Difference in time of arrival
$C(n, 2)$	Possible pairs
$L_{\Delta T_{ref}}$	Iso-ΔT curve
Ω_P	Plate domain
ϵ	Tolerable deviation in ΔT
KN	Closest neighbor
$\mathbf{X}$	Random input vecrtor
χ	Features space

$\mathbb{R}^P$	P dimensional Euclidean space
$\mathbf{Y}$	Response variable
$\mathbb{E}$	Expected value
$m(x)$	True regression function
D_n	Training set
m_n	Estimation of the Regression Function
Θ	Independent random variable
$\mathbb{1}$	Indicator Function
$\mathbb{E}$	Expectation respect to Θ
$D_n^*(\Theta_j)$	set of data points selected
$N_n(x;\Theta_j,D_n)$	number of point
$A_n(x;\Theta_j,D_n)$	cell containing x
e	Neper number
$S(x)$	Sigmoid
S_{lower}	Lower limit sigmoid
S_{upper}	Upper limit sigmoid
$\mathbf{y}$	data vector
q	percentile
$P_q(\mathbf{y})$	q-th percentile
inf	Lower limit
$\mathbb{R}$	Set of real numbers
$E_{90\|95}$	Radius associate to 95% confidence in localization
$f(n)$	Discrete signal
$S(n)$	Approximation functions
$d(n)$	Detail function
$h(n)$	FIR low-pass filter
$g(n)$	FIR high-pass filter
f_n	Frequency limit
$\psi(t)$	Mother wavelet
$g(t)$	Time dependent signal
γ	Wavelet coefficient
T	DWT threshold
$\mathcal{H}$	Hilbert transform
π	Pi number
a_n	Arbitrary complex scalars
c	Constant value
$sgn()$	Sign function
$\mathcal{F}$	Fourier Transform
$s_{USB}(t)$	Upper sideband signal
$s_{LSB}(t)$	Lower sideband signal
$g_a(t)$	Complex analytical signal
$E(t)$	Hilbert function envelope
L	Maximum likelihood estimate
σ	Variance
μ	Mean value
t	Time vector
$\zeta_{a,b}(x)$	CWT mother wavelet
t	Time vector

ζ_{FT}	Fourier transform of ζ
C_ζ	Condition of admissibility parameter
d	Distance
(x_c, y_c)	Centroid coordinates

Indexes

X	Abscissa axis
Y	Ordinate axis
i	Natural number

Contents

Introduction

In the global scenario, the aerospace industry is growing rapidly, in particular the commercial flights branch. It is estimated that global air traffic will double by 2035 (source: AIRBUS Global Market Forecast 2016-2035), along with earnings, the costs relating to the management and maintenance of aircraft will also increase proportionally. The focus is particularly on maintenance, repair and overhaul (MRO) costs which represent around $10 - 15\%$ of an airline's operating costs. In an attempt to reduce these costs and downtimes, systems are being developed for the continuous monitoring of structural health which are called Structural Health Monitoring systems.
The Structural Health Monitoring (SHM) is a system, which detect the various kind of damage and failure induced in the structure, also interpretation and assessment of the damage induced in the engineering structure in order to improve its reliability[1]. Simply integration of a non-destructive evaluation system into a structure for damage and failure prediction in the structure is called the SHM.
Structural Health Monitoring (SHM) aims to give, at every moment during the life of a structure, a diagnosis of the "state" of the constituent materials, of the different parts, and of the full assembly of these parts constituting the structure as a whole [2]. The state of the structure must remain in the domain specified in the design, although this can be altered by normal aging due to usage, by the action of the environment, and by accidental events. Thanks to the time-dimension of monitoring, which makes it possible to consider the full history database of the structure, and with the help of usage monitoring, it can also provide a prognosis (evolution of damage, residual life,etc.). It involves the integration of sensors, possibly smart materials, data transmission, computational power, and processing ability inside the structures. It makes it possible to reconsider the design of the structure as a part of wider systems.

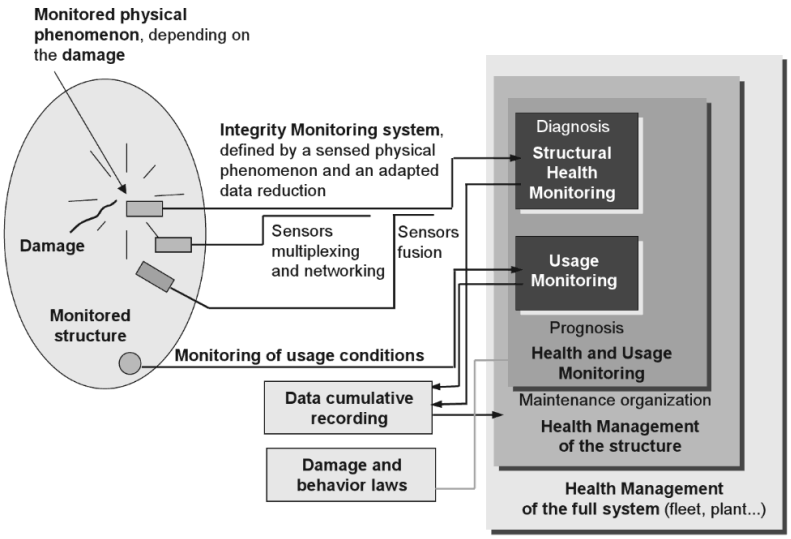


Figure 1: Principle and organization of a SHM system

In Figure-1, the organization of a typical SHM system is given in detail. The first part of the system, which corresponds to the structural integrity monitoring func-

tion, can be defined by:

- the type of physical phenomenon, closely related to the damage, which is monitored by the sensor,
- the type of physical phenomenon that is used by the sensor to produce a signal (generally electric) sent to the acquisition and storage sub-system.

Several sensors of the same type, constituting a network, can be multiplexed and their data merged. The signal delivered by the integrity monitoring sub-system, in parallel with the previously registered data, is used by the controller to create a diagnostic. Mixing the information of the integrity monitoring sub-system with that of the usage monitoring sub-system and with the knowledge based on damage mechanics and behavior laws makes it possible to determine the prognosis (residual life) and the health management of the structure (organization of maintenance, repair operations, etc.). Finally, similar structure management systems related to other structures which constitute a type of super system (a fleet of aircraft, a group of power stations, etc.) make possible the health management of the super system. Workable systems can be set up even if they are not as comprehensive as described here.
SHM, like Non-Destructive Evaluation (NDE), can be passive or active [3].
If the experimenter has equipped the structure with both sensors and actuators, the second one can generate perturbations in the structure, and then, the sensors monitor the response of the structure. In such a case, the action of the experimenter is "active monitoring". If the experimenter is just monitoring this evolution thanks to the embedded sensors, we can call his action "passive monitoring". For SHM, this sort of situation is encountered with Acoustic Emission (AE) techniques.
Through the use of AE technique, structures can be monitored in real time given that signals are emitted immediately when a crack occurs. These signals can also be analyzed in real time using to generate information regarding the nature of the crack or energy source [4]. AE techniques will be the main topic on which the thesis work is based, expecially its implementations and uses for impact localization over plate-like structures.
In particular, the following chapters will have the task of highlighting:

- Chapter 1 - Acoustic Emission: defines acoustic emission and takes up in depth the concepts relating to AE techniques which will then be implemented later.
- Chapter 2 - Probability of Localization: the theoretical aspects related to the probability of localization will be discussed, together with the methodology with which it will be applied to the results.
- Chapter 3 - Signal Processing and Features Extraction: the filtering methods of the acquired signal are exposed and subsequently how the key information for the localization of the AE will be extracted.
- Chapter 4 - Experimental Campaign: the experimental setup is introduced with the test articles used and the hardware and software instrumentation to acquire the AE signals
- Chapter 5 - Aluminum Plate Analysis: clarifies the methodology utilized for the localization of impact sources on the aluminum plate, in particular the

signal filtering techniques, TOA calculation and implementation of the ΔT method

- Chapter 6 - CFRP Plate Analysis: the studies carried out on the aluminum plate are repeated on the CFRP plate, highlighting the differences between the two analyses. In particular, the novelty of the sensor pairs' regions of influence will be exposed and how it combines with the classic ΔT method. Subsequently, the developed techniques will also be tested on H-S impact sources. The chapter concludes with the testing of machine learning localization algorithms and the comparison between the two types of approaches.
- Chapter 7 - Futures Perspectives: a second CFRP test article will be preset. The same experiments as on the first have been performed, but it is not analyzed yet.

1 Acoustic Emission

The definition of AE as given by ASTM E1316 (2014) is "The class of phenomena whereby transient elastic waves are generated by the rapid release of energy from a localized source or sources within a material, or the transient wave(s) so generated". Acoustic emission or stress wave emission describes the acoustic stress waves that emerged as a result of a rapid release of energy because of the microstructural changes that occur in materials[5]. Starting since 1950 with the work of Joseph Kaiser, researchers investigated the essentials of AE, established instrumentation particularly for AE, and considered the AE behavior of numerous resources [6]. AE was starting to be recognized for its exclusive competences as a NDT technique intended for checking dynamic procedures. Two types of analysis can be considered by AE. One is under time domain waveform which is related to basic parameter in a time domain of testing and the second is frequency domain waveform which considers the signal parameters with upcoming recorded frequency under testing. The extensively used signal measurement parameters in AE signal examination are counts, duration, amplitude, rise time, and the measured area under the corrected signal envelope that is also called relative energy [7].

- Amplitude: It is represented usually with A and shows the highest peak of the voltage signal. This is identical significant characteristic since it openly controls how much the AE event is detectable. Amplitude of the AE signal is immediately related to the magnitude of the source and varies over a wide range from micro volts to volts. According to Unnþórsson [9], the amplitudes of AE are usually mentioned in decibel scales.
- Counts: commonly are presented by N. When the amplitude of the signal is larger than the threshold, the number of pulses emitted by measurement circuitry is defined as N. Counts strongly rely on the AE properties and reverberant nature of the sensor and specimen material.
- Hits: when a signal exceeds the threshold, consequently a system channel accumulates data, this signal is known as the hit and describes an AE event. The number of events or hits per time determines the event rate. Both number of hits and number of counts determines the quantity of an AE activity.
- Duration: The time interval between a signal trigger and the time it reduces below the threshold value is known as the duration of the signal.
- MARSE Energy: Measured Area of the Rectified Signal Envelope is known as MARSE and sometimes represented by E and referred to as energy counts. This area is the zone below the envelope of the signal which is rectified and measured from the sensor. Energy is preferred over counts since it is sensitive to amplitude as well as duration, and it relies fewer on the threshold location and functioning frequency. Absolute (true) energy also is resulting from squared voltage signal divided by reference resistance of 10 $k\Omega$ over the time duration of acoustic emission waveform packet. This parameter indicates the true energy of an AE event from transient signals or of certain data rate interval of continuous AE signals. The absolute energy of a detected AE burst signal commonly expressed in atto-joules ($1\ aJ = 10^{-18}\ J$) and values can be

mentioned in logarithmic scales (dB, decibels). The energy is defined as

$$Et = \frac{1}{2}\int_{t1}^{t2} \oint_{+}^{2}(t)dt - \frac{1}{2}\int_{t1}^{t2} \oint_{-}^{2}(t)dt \tag{1}$$

- Rise time: is presented by R and defines the time interval between the maximum amplitude of the burst signal and the first threshold crossing. This parameter is frequently used when time dependent processes such as vibration or dynamic loading are involved.
- Threshold: is a kind of set up parameter which is used for the elimination of electronic background noises. Its main objective is to eliminate as much background noise as it can. However, a balance should be struck so that the threshold does not also eliminate signals that are weak but useful given that background noises are characterized by low amplitude.

AE are widely used in the SHM field, in the course of the discussion we will focus in particular on AE localization techniques for Two-Dimensional (2D) Structures [10]. Their purpose is to define an area within which the presence of the source is estimated so as to be able to limit the area to be inspected for any damage.
Among the most used methods we find triangulation method, beam forming, ΔT method and machine learning.
The triangulation method uses the time difference between the sound wave and different sensors to locate the sound source. The triangulation method has the following requirements: (1) wave velocity is constant in the medium; (2) arrival of the wave signal can be caught perfectly; (3) a straight wave path; (4) sensor locations are accurately determined; and (5) it requires a lot of sensors. Recent studies conducted by Kanakaraj [11] report a high accuracy in the localization of AE inside a concrete block, reaching an absolute error on centimeter level. The constraint on the isotropy of the material remains.
Going further, the beamforming method was first proposed by McLaskey for sound source localization [12]. The core principle is that the transmitting end will weight the data to be sent and then form a focused narrow beam, which is like a laser beam and directly points towards the target receiving user. The main purpose of doing so is to enhance the signal strength sent to the target user, reduce signal attenuation during propagation and significantly improve the demodulation signal-to-noise ratio of the target user, especially for users located at the edge of the cell with relatively weak signals. The beam forming method has been successfully applied in multiple fields such as sonar, telecommunications, radar, sound source recognition and localization. In the most recent studies we can find those conducted by Tai [13], proposed a method to determine the position of a sound source in an anisotropic structure by using an L-shaped transducer array. The test was carried out on a homogeneous thin-walled steel plate with the size $700 \times 700 \times 5\ mm$. The accuracy achieved through the use of the fast Bartlett beamforming algorithm is around 2 mm on X and Y. The limitations are (1) that it requires a large number of sensor groups; (2) when comparing the shape of signals in beam forming, all sensors in the array should have similar amplitude and phase responses; and (3) to avoid spatial aliasing, the beam forming array needs to have a relatively small aperture.
Regarding the last two methods mentioned, ΔT method and machine learning, they

are widely used in the aerospace sector for the localization of AE, thanks to their high precision and reliability. In the next sections we will go into detail about the methods, their implementation and motivations.

1.1 ΔT method

Before examining the methodology it is necessary to clarify what the term ΔT refers to. By ΔT we refer to the difference in wave time of arrival (TOA) between pairs of different sensors. The Time of Arrival is defined by Prof. Christian Grosse as the absolute time when particular phase of an acoustic emission signal signal (usually the very first wave onset) arrives at a sensor. Corrections could be required for the sensor/instrument death time (caused by electronic delays) and also for noise (random vibrations caused by other sources) influencing the accuracy of onset detection. It can be calculated in various ways as we will see later, for example using threshold trespass, Maeda's Akaike Information Criteria or Continuous Wavelet Transform. Having laid the foundations, it is now possible to go into the details of the methodology implementation. There are five steps associated with this location method, which are detailed below [14]:

- Determine area of interest: The ΔT source location method is not envisaged as a method for source location for complete coverage of a part or structure. Though this would be possible, it would be time consuming and would not be justified by the improvement in location. It is suggested that ΔT source location is used as a tool to improve source location around important/complicated areas, including areas where high stress levels are predicted. Therefore, the first step in the ΔT location process is to determine areas of interest on the structure. A further advantage of this method of source location is that new areas of interest can be added during or post test and therefore active areas identified by TOA can also be more accurately assessed using the ΔT source location method, supporting the proposal to run other types of location in parallel.

- Construct grid: a grid is constructed on the component in which future AE events will be located; the higher the resolution of the grid, the greater the accuracy of this method. Higher resolution, up to a point, will result in better location. From basic wave theory it is only possible to locate to within 1 sensor diameter of the source. It is important to note that the grid does not have to be in the same orientation as the sensors monitoring it. Sources are located with reference to the grid applied to the structure and not the sensor location. The sensor location is not required to locate the source.

- Conduct artificial source events to obtain time of arrival data: artificial source events are conducted at nodes in the grid to provide the AE time of arrival at each sensor; the time of occurrence of the source is not required. Several events at each node are required to provide an average result and to allow erroneous data to be removed from the data set. Practical work has shown that at least five events at each node are required to provide reliable data. It is not essential to have event data from every point in the grid and the grid density can be increased in complex areas if required. Missing data points can

be interpolated from the nodes in the surrounding area.

- Calculate ΔT maps: For each artificial source event, a difference in time of arrival or "ΔT" can be calculated for each sensor pair. Having a number n of sensors available, the possible pairs will be:

$$C(n,2) = \binom{n}{2} = \frac{n(n-1)}{2} \tag{2}$$

An average ΔT is calculated for each grid node. This information is then interpreted to develop the grid or map for each sensor pair. These maps display contour lines of equal ΔT for each pair of sensors. Interpolation can be used to increase the density the grid.

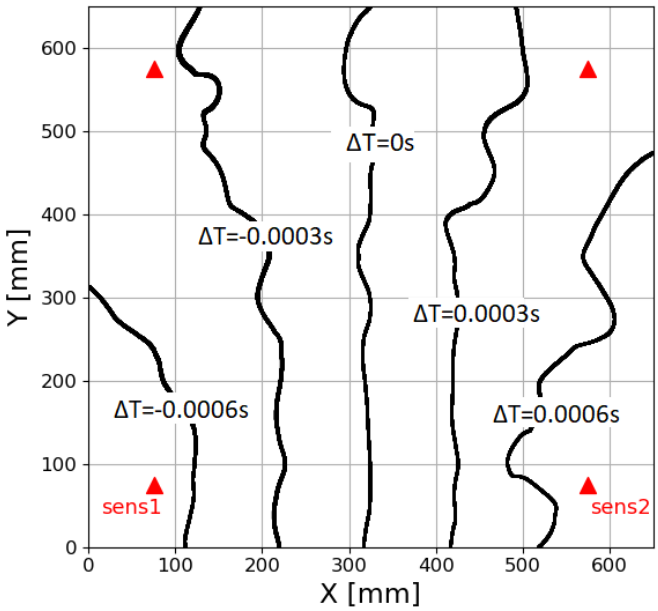


Figure 2: Iso-ΔT curves for sensors pair 1-2 associate to different ΔT

- Compare actual data : By calculating the ΔT for each sensor pair from an actual AE event, a line can be constructed on each ΔT map displaying possible source locations. The line referring to a certain ΔT is defined as the locus of the points to which a ΔT equal to the reference one corresponds to less than an ϵ deviation.

$$L_{\Delta T_{ref}} = \{P(x,y) \in \Omega_P \mid |\Delta T(x,y) - \Delta T_{ref}| \leq \epsilon\} \tag{3}$$

with $L_{\Delta T_{ref}}$ is the line (or region) that represents the locus of the points associated with the reference ΔT, $P(x,y) \in \Omega$ indicates a generic point (P) with coordinates (x,y) belonging to the plate domain (Ω_P), $\Delta T(x,y)$ is the ΔT calculated for that specific point in space, ΔT_{ref} is the reference ΔT (usually the one measured experimentally between the pair of sensors) and ϵ is the maximum tolerated deviation.
By overlaying results from each of the sensor pairs, a convergence point is identified, indicating the source location. A minimum of three sensors are

required to provide a point location. Using more sensors will increase the confidence of location estimation. Theoretically all lines will cross at one location, however, as with all practical tests, the reality is that not all lines will cross at the same point.

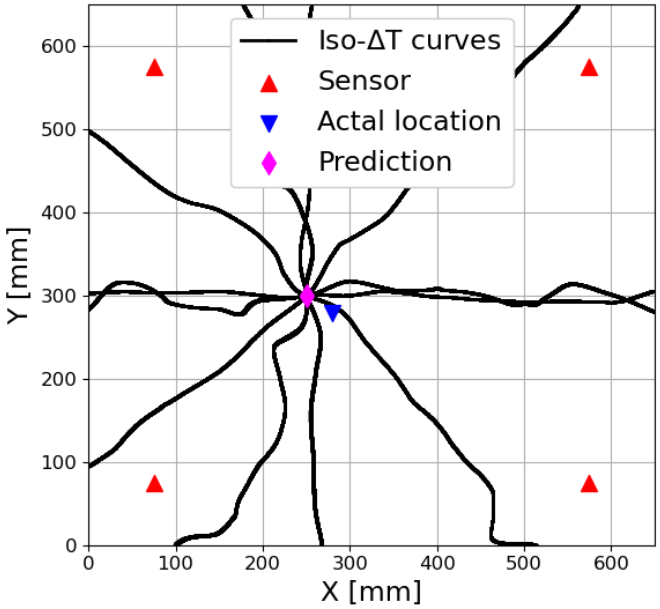


Figure 3: AE localization using iso-ΔT curves intersection

The advantage of this method is it involves no assumptions regarding wavepaths and wavespeeds nor any in-depth theoretical calculations. The studies conducted by Al-Jumaili and Safaa Kh. [16] in 2016 represent a valid example of implementation. They conducted two tests. The first was conducted on a simple geometry specimen made from 20 mm thick ASTM 516 gr 70 steel with overall dimensions of 90 $mm \times 2$ m, within this specimen two geometric features were present; the first was 4.4 mm wide and 10 mm deep v-notch and the other was a 20 mm diameter half circle cut. The second test was conducted on an aerospace grade 2024-T3 aluminium plate, with dimensions of 370 $\times$ 200 mm with a thickness of 3.18 mm. The specimen contained a series of differing diameter circular holes. On both test articles, 10 H-S sources were produced on 10 different points, the localization accuracy on the first specimen reached 3.13 mm while on the second it stopped at 3.88 mm.

1.2 Machine learning

The term machine learning refers to the subset of artificial intelligence (AI) based on algorithms capable of "learning" patterns from training data and subsequently making accurate inferences on new data [17]. This pattern recognition capability allows machine learning models to make decisions or predictions without explicit, coded instructions.
This algorithm involves a target or result variable, also known as the dependent variable, which may be predicted based on a specific collection of predictors, also known as independent variables. With this set of variables, we construct a function that maps inputs to the intended output. The training procedure persists until the model attains the desired degree of accuracy in the training data. The supervised learning techniques commonly used for regression analysis include decision tree, random forest, closest neighbor (KN), and logistic regression.
Unsupervised learning is used when there is data accessible solely in the input form and no corresponding output variable. These algorithms utilize statistical models to analyze the fundamental patterns inside the data in order to gain a deeper understanding of its characteristics. Clustering is a prominent category of unsupervised algorithms. This technique involves the identification of intrinsic groupings within the data, which are then utilized to forecast the output of hidden inputs.
Reinforcement learning is utilized when the objective is to make a sequence of choices that lead to a final reward. Throughout the learning process, the artificial worker is given either rewards or penalties based on the acts it carries out. The objective is to optimize the overall reward. Examinations encompass the acquisition of knowledge related to playing computer games or carrying out activities involving robots, with the overarching objective. By employing this method, the machine undergoes training to produce precise determinations. An instance involves subjecting the device to a setting where it consistently improves its performance by undergoing repeated trials and adjustments [21].
These types of algorithms have only recently been used in SHM and in particular in AE localization, but they are finding widespread use in industry thanks to the advantage of not requiring knowledge of the wave speed and therefore being able to be implemented on composite structures. Its disadvantage is that it requires a large amount of data and high computational requirements [18]. In the literature it is possible to find various examples of implementation both on articles made of isotropic and composite materials. Referring to the experiments conducted by Kim in 2014 [19], on both aluminum and composite plates of variable dimensions ($900 \times 900\ mm^2$ and $600 \times 600\ mm^2$) an error between 9.8 mm and 1.9 mm was obtained. We find further confirmation in T.Fu's studies [20], which he used a backpropagation artificial neural network for AE source localization from the data collected by a prismatically arranged sensor matrix on an anisotropic carbon fiber polymer plate ($300\ mm \times 300\ mm \times 1.8\ mm$).The average error encountered was only 4.3 mm, and the test point of the maximum error was 6.4 mm.
In this chapter and subsequent applications we will focus mainly on random forest algorithms. Below are reported a brief but precise explanations of the random trees, which represent the fundamental constituents of the random forest, and then of the random forest itself.

1.2.1 Decision Trees

Decision Trees are a technique utilized in the fields of Statistics, data mining, and machine learning. It falls within the category of supervised machine learning. Data analysis technology categorizes data into different entities that are potentially associated with a specific procedure. There are two categories of such entities: contract and paper.
The supervised learning approach employs the decision tree as a prediction model to examine the observations of an item in the branches and deduce the target value of the item in the leaves. The decision nodes symbolize the segmentation of data, while the sheets symbolize the outcomes.
The following is the operational process of the terminal for your attention.

- The root node: the starting point of the decision tree. The root node symbolizes the complete dataset, which is partitioned into two or more groups that can be compared.
- Leaf node: correspond to the ultimate result.

The algorithm is unable to further divide the tree once it reaches the leaf node. Splitting refers to the process of separating the root node or decision node into distinct sub-nodes based on specific parameters. Pruning is the act of removing superfluous branches from a tree. By eliminating irrelevant branches, one can reach a conclusion much more quickly.
The parent node is the root node of the tree, while the other nodes are its children. A subtree is created when the primary tree is divided, resulting in new subtrees and branches.

Machine learning encompasses two primary categories of decision trees, which are distinguished by the goal variable:

1. Classification Trees: specific kind of decision tree that can handle discrete values for the target variable. The decision tree is based on a categorical variable, with the possible value being "yes" at the first position. It eliminates the outcome following the evaluation of the provided data. A classification or judgment tree is an example of a tree structure that has two or more branches, each dependent on a distinct set of values
2. Regression Trees: specific kind of Decision Tree that can handle target variables with continuous values. The decision tree is designed to handle continuous variables, where the values can be represented as real numbers. Typically, the creation of a regression tree includes utilizing inputs that consist of a mixture of continuous and discrete variables. Every decision node examines the input to evaluate the value of the variable. The regression tree utilizes a binary recursive division method. At each iteration, the data is divided into parts, which are then further divided into smaller groups when ascending the branch.

1.2.2 Random Forest

Random forests can be used either for a categorical response variable, such as "classification", or a continuous response, referred to as "regression". Similarly, expectation

variables can be either categorical or continuous. From a computational point of view, random forests are attractive because they:

- Naturally cope with both regression and classification (multilayer),
- Relatively quick to train and predict,
- Only depend on one or two adjusting parameters,
- Have a built-in Estimate of the generalization error,
- Can be used directly for high dimensional problems,
- Can be easily implemented in parallel.

Random forest is a machine learning technique that combines numerous decision trees to reduce the correlation among feature data. Simultaneously, the computational cost of RF is $O(n)$ (n represents the number of samples) when dealing with huge amounts of data. Additionally, the technique can be executed in parallel due to this integration, resulting in increased speed [21].

Now a concise mathematical presentation of the algorithm for building a random forest is provided.
The general framework is nonparametric regression estimation, in which an input random vector $\mathbf{X} \in \chi \subset \mathbb{R}^P$ is observed, and the goal is to predict the square integrable random response $\mathbf{Y} \in \mathbb{R}$ by estimating the regression function $m(x) = \mathbb{E}[\mathbf{Y}|\mathbf{X} = x]$. With this aim in mind, we assume we are given a training sample $D_n = ((X_1, Y_1), ..., (X_n, Y_n))$ of independent random variables distributed as the independent prototype pair(X, Y). The goal is to use the data set D_n to construct an estimate $m_n : \chi \to \mathbb{R}$ of the function m. In this respect, we say that the regression function estimate m_n is (mean squared error) consistent if $\mathbb{E}[m_n(\mathbf{X}) - m(\mathbf{X})]^2 \to 0 \quad as \quad n \to \infty$ (the expectation is evaluated over $\mathbf{X}$ and the sample D_n).

A random forest is a predictor consisting of a collection of M randomized regression trees. For the j-th tree in the family, the predicted value at the query point x is denoted by $m_n(x; \Theta_j, D_n)$, where $\Theta_1, ..., \Theta_M$ are independent random variables, distributed the same as a generic random variable Θ and independent of D_n. In practice, the variable Θ is used to resample the training set prior to the growing of individual trees and to select the successive directions for splitting. In mathematical terms, the j-th tree estimate takes the form

$$m_n(x; \Theta_j, D_n) = \sum_{i \in D_n^*(\Theta_j)} \frac{\mathbb{1}_{x_i \in A_n(x; \Theta_j, D_n)} Y_i}{N_n(x; \Theta_j, D_n)} \tag{4}$$

where $D_n^*(\Theta_j)$ is the set of data points selected prior to the tree construction, $A_n(x; \Theta_j, D_n)$ is the cell containing x, and $N_n(x; \Theta_j, D_n)$ is the number of (preselected) points that fall into $A_n(x; \Theta_j, D_n)$.

At this stage, we note that the trees are combined to form the (finite) forest estimate:

$$m_{M,n}(x; \Theta_1, ..., \Theta_M, D_n) = \frac{1}{M} \sum_{j=1}^{M} m_n(x; \Theta_j, D_n) \tag{5}$$

Since M may be chosen arbitrarily large (limited only by available computing resources), it makes sense, from a modeling point of view, to let M tends to infinity, and consider instead of the (infinite) forest estimate:

$$m_{\infty,n}(x, D_n) = \mathbb{E}_\Theta[m_n(x; \Theta, D_n)] \tag{6}$$

$\mathbb{E}_\Theta$ denotes the expectation with respect to the random parameter Θ, conditional on D_n. In fact, the operation "$M \to \infty$" is justified by the law of large numbers, which asserts that almost surely, conditional on D_n,

$$\lim_{M \to \infty} m_{M,n}(x; \Theta_1, ..., \Theta_M, D_n) = m_{\infty,n}(x; D_n) \tag{7}$$

[22].

2 Probability of Localization

The Probability of Localization (POL) is calculated based on the method of the classical Probability of Detection (POD) approach derived from MILHDBK1823 [39].
In the following sections the POL is referred as the confidence in AE localization associated to a certain width of the localization area. The intent is to find the radius $E_{90|95}$ that identifies an area in which the impact occurs with a 95% confidence for 90% of tests.
The steps to compute the POL and $E_{90|95}$ are listed below.

(1) Logistic curve definition: after obtaining the localization error values for the test points, is possible to extract thought regression the parameters x_0 and k of the sigmoid that best fits the data trend.

$$S(x) = \frac{1}{1 + e^{-k(x-x_0)}} \tag{8}$$

defining:

- x: localization error (or radius)
- k: slope
- x_0: median value of test localization error

By definition the sigmoid has asymptote at 0 and 1 therefore it needs to be scaled between 0% and 100% (percentage of tests).

(2) Bootstrap: a random number of values are selected M times from the localization error vector. For each set of random values the sigmoid is constructed (step (1)) and the results are saved in the columns of an $n \times M$ matrix, n is the number of points describing the curve. In this way we simulate the real variability that would occur when carrying out various tests.

(3) Bootstrap percentile interval definition: the average sigmoid is extracted from the matrix as the average of the values along the rows and represents the center of the confidence interval. The sigmoids representing the upper and lower limits are the curves associated with the 2.5 (lower) and 97.5 (upper) percentiles calculated starting from the $n \times M$ matrix.

$$\begin{aligned} S_{lower} &= P_{2.5}(\mathbf{Y}) \\ S_{upper} &= P_{97.5}(\mathbf{Y}) \end{aligned} \tag{9}$$

Defining S_{lower} and S_{upper} as limit sigmoid, $\mathbf{Y}$ as the $n \times M$ matrix, $P_{2.5}$ and $P_{97.5}$ indicate the statistical operator of the 2.5^{th} and 97.5^{th} percentile:

$$P_q(\mathbf{y}) = \inf \left\{ y \in \mathbb{R} \mid \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{\{y_i \leq y\}} \geq \frac{q}{100} \right\} \tag{10}$$

Now the average sigmoid and the two limits have been obtained. To adapt them to the real problem, the asymptotes at $y = 0\%$ and $y = 100\%$ are eliminated, forcing the curve to take on a value of $x = 0$ at $y = 0\%$ and a finite value (depends on the maximum value in the error dataset) at $y = 100\%$.

Through this procedure we obtain an interval in which other possible sigmoids associated with new experimental datasets fall with a probability of 95%.

(4) $E_{90|95}$ definition: the radius $E_{90|95}$ sought represents the value assumed by the upper limit sigmoid S_{upper} at $y = 90\%$.

3 Signal Processing and Features Extraction

3.1 Signal Processing

Processing of Acoustic Emission signal is an essential operational phase in Structural Health Monitoring, necessary to transform the raw data acquired by the sensors into usable diagnostic information. In aeronautical contexts, the signals originating from an impact on a CFRP plate are inevitably superimposed on various sources of disturbance such as low frequency mechanical vibrations or electromagnetic noise induced by the instrumentation itself.
Noise filtering constitutes the first and most delicate step of post-processing. Its aim is to maximize the Signal-to-Noise Ratio while rigorously preserving the physical characteristics of the transient wave (such as arrival time, peak amplitude, rise time and spectral content), which are the strictly necessary input parameters for localization algorithms.

In this chapter we examine moving average filter, discrete wavelet filter and Hilbert transform filter that will be then implemented in subsequent analyses.

3.1.1 Moving Average Filter

Moving Average is the most common used in DSP, even if is one the most easy to implement it is is optimal for reducing random high frequency noise while retaining a sharp step response [23].
As suggested by the name, the filter operate by averaging a fixed number of point of the input signal, that constitute the averaging window, to produce each point in the output signal:

$$y[i] = \frac{1}{M} \cdot \sum_{j=0}^{M-1} x[i-j] \tag{11}$$

expressing $y[i]$ as the filter output point corresponding to i-th time instant, M is the size of averaging window, x[ ] are the value of raw input signal.

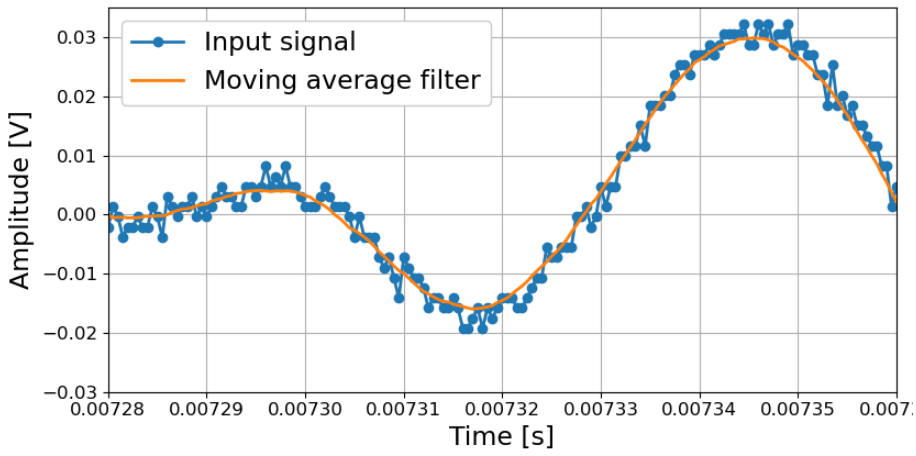


Figure 4: MAF 20 point symmetrical window

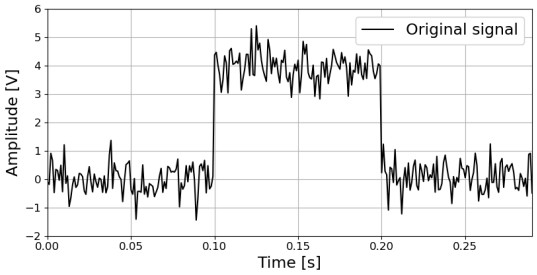
The average window can be dynamically positioned around the time instant, for example is possible to position the window on one side, giving importance to what happens only before or after the i-th time:

$$y[80] = \frac{x[80] + x[81] + x[82] + x[83]}{4} \tag{12}$$

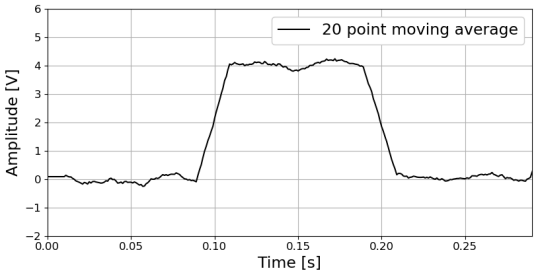
or symmetrically with half of the points on both side:

$$y[82] = \frac{x[80] + x[81] + x[82] + x[83] + x[84]}{5} \tag{13}$$

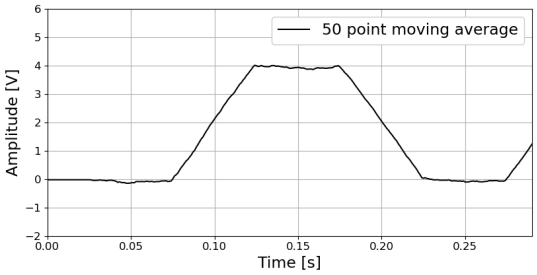
In this way the output point is equally influenced by what happened before and after over time, but due to the finitude of the signal using a symmetrical window fail on the edge of the input signal because whether we are at the beginning or the end we will lack one or more point for the average process, that's why a dynamically positioned window may prove to be the best approach.



(a) rectangular pulse with random noise



(b) MAF 20 points window



(c) MAF 50 points window

Figure 5: Example of different window size MAFs applied over a rectangular pulse with random noise

Let's now delve deeper into the issue of noise reduction while keeping the sharpest step response.
In panel (a), the signal is a pulse obscured by random noise. Panels (b) and (c) show how applying a moving average filter smooths the signal: it successfully reduces the noise amplitude (a positive effect), but also blurs the signal's sharp edges (a negative trade-off). Among all possible linear filters, the moving average offers the lowest noise level for a given edge sharpness.
The degree of noise reduction is proportional to the square root of the number of points used in the average. For instance, using a 100-point moving average filter

results in a tenfold noise reduction.
Suppose we want to design a filter with a fixed edge sharpness, defined by a step response rising over eleven points. This constrains the filter kernel to have exactly eleven coefficients. The key optimization question becomes: how should we choose these eleven values to minimize output noise?
Because the noise is random and uniformly distributed across all input points, no sample holds more importance than another. Assigning higher weight to any single input would not improve noise reduction. Therefore, the lowest noise is achieved by treating all input samples equally, as the moving average does. This makes it the premier filter for time domain encoded signals.
However, the moving average is the worst filter for frequency domain encoded signals, with little ability to separate one band of frequencies from another. Relatives of the moving average filter include the Gaussian, Blackman, and multiplepass moving average. These have slightly better performance in the frequency domain, at the expense of increased computation time.

3.1.2 Discrete Wavelet Filter

The wavelet transform represents an alternative approach to the Fourier transform, instead of employing infinite-duration sine and cosine functions, the wavelet transform utilizes waveforms of finite duration, known as wavelets [35]. A wavelet is characterized by a limited temporal support, zero mean, and typically exhibits irregular and asymmetric behavior.
Wavelet analysis is based on a windowing technique with variable-sized regions, enabling long time windows for accurate low-frequency analysis and short time windows for high-frequency resolution. This property is referred to as multiresolution.
Compared to the Fourier transform, the wavelet transform provides the ability to perform localized signal analysis, detect trends and discontinuities in higher-order derivatives, and allow signal compression and denoising with minimal energy loss.
The discrete wavelet transform (DWT), as introduced by Mallat (1989), decomposes a discrete signal, $f(n)$ into a hierarchical set of orthogonal components: approximation functions, $S(n)$ and detail functions $d(n)$.
The approximation components, corresponding to the high-scale (low-frequency) part of the signal, are obtained through a finite impulse response (FIR) low-pass filter $h(n)$, while the detail components, representing low-scale (high-frequency) information, are extracted using an FIR high-pass filter $g(n)$.
At each decomposition level, the signal is passed through the two complementary filters, resulting in two sub-signals: $S_1(n)$ and $d_1(n)$. The approximation $S_1(n)$ is further decomposed into $S_2(n)$ and $d_2(n)$, and the process continues iteratively up to the desired level [24].

Level	Frequencies	Samples
3	0 to $f_n/8$	4
3	$f_n/8$ to $f_n/4$	4
2	$f_n/4$ to $f_n/2$	8
1	$f_n/2$ to f_n	16

Table 2: Frequency range associated with level decomposition

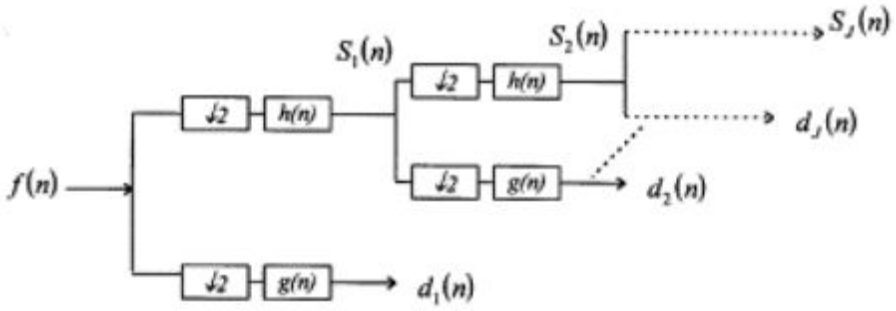


Figure 6: Signal decomposition using DWT at different levels

To avoid redundancy and limit the number of samples, the decomposition includes a downsampling step, which retains every other sample. This operation introduces spectral overlap between approximation and detail components, which can be mitigated by selecting appropriate filters for both the decomposition and reconstruction stages.

At each level in the above diagram the signal is decomposed into low and high frequencies. Due to the decomposition process the input signal must be a multiple of 2^n where n is the number of levels.

For example a signal with 32 samples, frequency range 0 to f_n and 3 levels of decomposition, 4 output scales are produced.

After each level of decomposition we will get a detail coefficient that describe the signal in a specified frequency range.

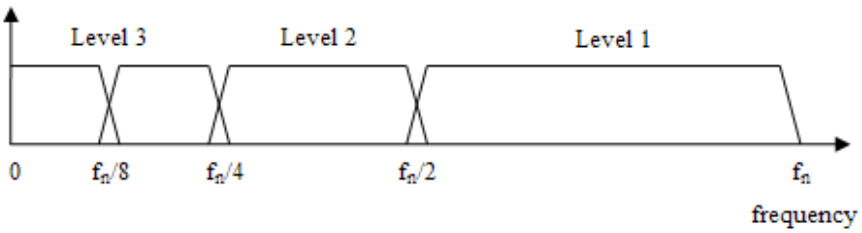


Figure 7: Frequency domain representation of the DWT

Reconstruction is achieved through an inverse process that includes upsampling—i.e., the insertion of zero-valued samples between data points—followed by filtering, allowing the original signal to be perfectly reconstructed from its components.

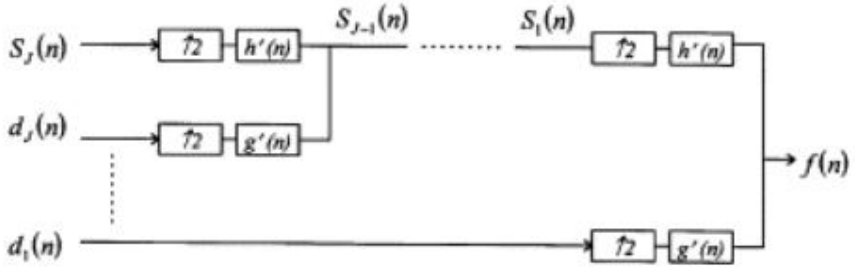


Figure 8: Signal reconstruction using DWT at different levels

The reconstruction involves using another set of low-pass, $h'(n)$, and high-pass, $g'(n)$, FIR filters. The choice of $h(n)$, $g(n)$, $h'(n)$, and $g'(n)$ defines the type of mother wavelet used as Haar or Daubechies wavelets.
Before exploring the features of different mother wavelets, Let's delve deeper into the connection between a mother wavelet and the detail coefficient obtained at level j.
In the case of the discrete wavelet transform, the mother wavelet $\psi(t)$ is shifted and scaled by powers of two:

$$\psi_{jk}(t) = \frac{1}{2^j}\psi\left(\frac{t - k2^j}{2^j}\right) \tag{14}$$

where j is the scale parameter and k is the shift parameter, both integer.
Defining the time dependent signal $g(t)$ of length 2^n, the wavelet coefficient γ of $g(t)$ is the projection of the signal onto a wavelet:

$$\gamma_{jk} = \int_{-\infty}^{+\infty} x(t) \cdot \psi_{jk}(t)\, dt \tag{15}$$

respect to equation (4):

$$\gamma_{jk} = \int_{-\infty}^{+\infty} x(t) \cdot \frac{1}{2^j}\psi\left(\frac{t - k2^j}{2^j}\right)\, dt \tag{16}$$

Fixing the scale j, the coefficient γ_k can be seen as as a convolution of $x(t)$ with a dilated, reflected, and normalized version of the mother wavelet $\psi_{jk}(t)$, sampled at the point 1, 2^j, $2 \cdot 2^j$, ..., 2^n, this is the precisely definition of the detail coefficients give at level j of the discrete wavelet transform.

Let's now analyze Daubechies wavelets features.
The Daubechies wavelets are a family of orthogonal wavelets defining a discrete wavelet transform and characterized by a maximal number of vanishing moments (certain quantitative measures related to the shape of the function's graph), A, for some given domain. With each wavelet type of this class, there is a scaling function (called the father wavelet) which generates an orthogonal multiresolution analysis.
In general the Daubechies wavelets are chosen to have the highest number A of vanishing moments, (this does not imply the best smoothness) for given support width (number of coefficients) 2A. There are two naming schemes in use, DN using the length (D2, D4,..), and dbA referring to the number of vanishing moments (db2, db4,...).

Each wavelet has a number of zero moments or vanishing moments equal to half the number of coefficients. For example, D2 has one vanishing moment, D4 has two, etc. A vanishing moment limits the wavelets ability to represent polynomial behaviour or information in a signal. For example, D2, with one vanishing moment, easily encodes polynomials of one coefficient, or constant signal components. D4 encodes polynomials with two coefficients, i.e. constant and linear signal components; and D6 encodes 3-polynomials, i.e. constant, linear and quadratic signal components. This ability to encode signals is nonetheless subject to the phenomenon of scale leakage, and the lack of shift-invariance, which arise from the discrete shifting operation during application of the transform.
For instance, the first four filter coefficients for the fourth-order Daubechies wavelet (Db4) referred to the previous graphs are given by:

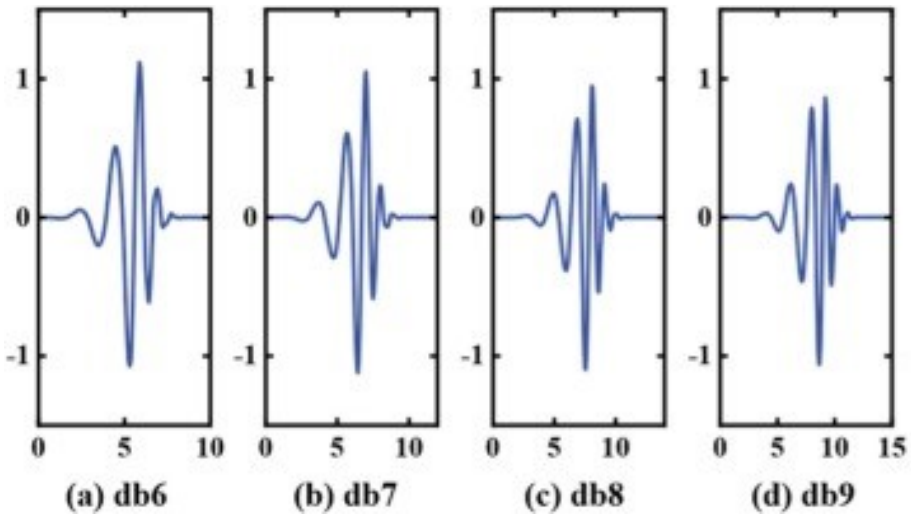


Figure 9: Db wavelets

$$h(n) = [-0.0106,\ 0.0329,\ 0.0308,\ -0.1870,\ -0.0280,\ 0.6309,\ 0.7148,\ 0.2304] \tag{17}$$

$$g(n) = [-0.2304,\ 0.7148,\ -0.6309,\ -0.0280,\ 0.1870,\ 0.0308,\ -0.0329,\ -0.0106] \tag{18}$$

$$h'(n) = [0.2304,\ 0.7148,\ 0.6309,\ -0.0280,\ -0.1870,\ 0.0380,\ 0.0329,\ -0.0106] \tag{19}$$

$$g'(n) = [-0.0106,\ -0.0329,\ 0.0308,\ 0.1870,\ -0.0208,\ -0.6309,\ 0.7148,\ -0.2304] \tag{20}$$

In signal analysis with the DWT is possible to determine the detail level or detail coefficient in a certain frequency range that contain most of the noise, using a soft thresholding function based on standard deviation of the coefficient is possible to reduce the noise in the signal:

$$\text{soft thresholding}(x, T) = \begin{cases} \text{sign}(x) \cdot (|x| - T), & \text{if } |x| > T \\ 0, & \text{otherwise} \end{cases} \tag{21}$$

where T is referred as threshold. Therefore the wavelet coefficients with an absolute value lower than the threshold T are deleted; the higher ones are reduced in module, maintaining the main data structure.

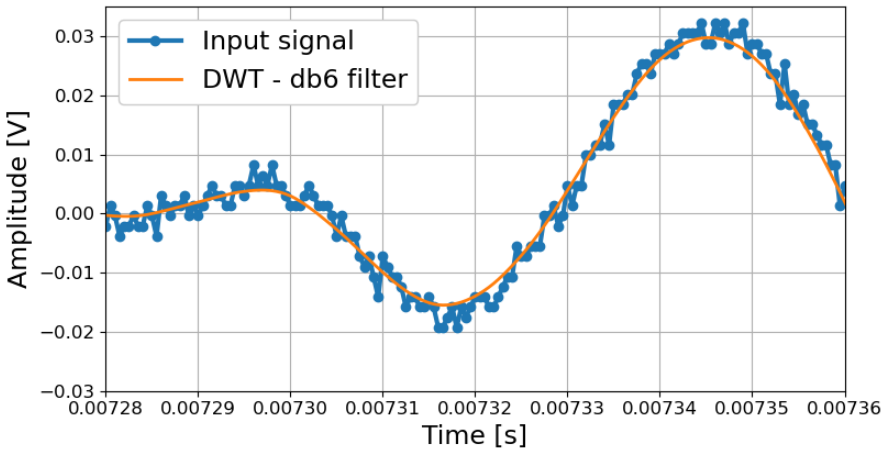


Figure 10: Denoising with db6 DWT

3.1.3 Hilbert Transform Filter

The Hilbert transform $\mathcal{H}[g(t)]$ of a signal $g(t)$ is defined as

$$\mathcal{H}[g(t)] = g(t) * \frac{1}{\pi t} = \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{g(\tau)}{t-\tau}\,d\tau = \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{g(t-\tau)}{\tau}\,d\tau. \tag{22}$$

The integral is improper due to a singularity and infinite limits [25]. Thus, the Hilbert transform is defined as the *Cauchy principal value*, whenever this value exists:

$$\mathcal{H}[g(t)] = \frac{1}{\pi}\lim_{\varepsilon\to 0^+}\left(\int_{t-1/\varepsilon}^{t-\varepsilon} \frac{g(\tau)}{t-\tau}\,d\tau + \int_{t+\varepsilon}^{t+1/\varepsilon} \frac{g(\tau)}{t-\tau}\,d\tau\right). \tag{23}$$

Cauchy principal value is obtained by considering a finite range of integration that is symmetric about the point of singularity, but which excludes a symmetric subinterval, taking the limit of the integral as the length of the interval approaches ∞ while, simultaneously, the length of the excluded interval approaches zero.

Let's analyze some basic property of Hilbert transform.
Clearly the Hilbert transform of a time-domain signal $g(t)$ is another time-domain signal $\hat{g}(t)$.

Linearity: if a_1 and a_2 are arbitrary (complex) scalars, and $g_1(t)$ and $g_2(t)$ are signals, then

$$\mathcal{H}[a_1 g_1(t) + a_2 g_2(t)] = a_1\hat{g}_1(t) + a_2\hat{g}_2(t). \tag{24}$$

Transform of a constant signal: for any constant value c

$$\mathcal{H}[c] = 0. \tag{25}$$

Time shifting and scaling:

$$\mathcal{H}[g(t-t_0)] = \hat{g}(t-t_0), \quad \mathcal{H}[g(at)] = \operatorname{sgn}(a)\hat{g}(at). \tag{26}$$

Convolution:

$$\mathcal{H}[g_1(t) * g_2(t)] = \hat{g}_1(t) * g_2(t) = g_1(t) * \hat{g}_2(t).$$

Time derivative: Hilbert transform of the derivative of a signal is the derivative of the Hilbert transform

$$\mathcal{H}\left[\frac{d}{dt}g(t)\right] = \frac{d}{dt}\mathcal{H}[g(t)]. \tag{27}$$

Interaction with the Fourier Transform: The signal $1/(\pi t)$ has Fourier transform

$$\mathcal{F}\left[\frac{1}{\pi t}\right] = -j\operatorname{sgn}(f), \tag{28}$$

where *sgn* is the sign function:

$$\operatorname{sgn}(f) = \begin{cases} 1 & \text{if } f > 0 \\ 0 & \text{if } f = 0 \\ -1 & \text{if } f < 0 \end{cases} \tag{29}$$

so $-j\operatorname{sgn}(f)$ is a piecewise-defined complex function:

$$-j\operatorname{sgn}(f) = \begin{cases} -j & \text{if } f > 0 \\ 0 & \text{if } f = 0 \\ j & \text{if } f < 0 \end{cases} \tag{30}$$

A Hilbert transform $\hat{g}(t)$ has as Fuorier transform:

$$\mathcal{F}[\hat{g}(t)] = \hat{G}(f) = -j\operatorname{sgn}(f)G(f). \tag{31}$$

That is, in the frequency domain, the Hilbert transform does not change the *magnitude* of $G(f)$, but it changes only the *phase*:

- For $f > 0$, it multiplies by $-j$, which corresponds to a phase shift of $-\pi/2$.
- For $f < 0$, it multiplies by $+j$, corresponding to a phase shift of $+\pi/2$.

The function $-j\operatorname{sgn}(f)$ represents an *ideal phase-shifting filter* that implements the Hilbert transform in the frequency domain; this property is particularly useful when used to generate the analytical signal, it allows the envelope to be calculated without introducing delays.

Energy spectral density: suppose that $g(t)$ is an energy signal, since

$$|\hat{G}(f)| = |G(f)|. \tag{32}$$

both $\hat{G}(f)$ and $G(f)$ have exactly the same energy spectral density. If $G(f)$ is bandlimited to [ ]Hz then so is $\hat{G}(f)$. It also follows that $\hat{g}(t)$ has *exactly the same*

energy as $g(t)$.

Symmetry: if $g(t)$ is real-valued, then $G(-f) = G^*(f)$, and $\hat{G}(-f) = \hat{G}^*(f)$. Recall that if $g(t)$ is even (so that $g(-t) = g(t)$) then $G(f)$ is purely real-valued while if $g(t)$ is odd (so that $g(-t) = -g(t)$) then $G(f)$ is purely imaginary-valued. Now if $G(f)$ is purely real-valued then certainly $\hat{G}(f)$ is purely imaginary-valued (and vice-versa). Thus if $g(t)$ is even, then $\hat{g}(t)$ is odd and if $g(t)$ is odd, then $\hat{g}(t)$ is even.

Orthogonality: if $g(t)$ is real-valued and square integrable:

$$\langle g(t), \hat{g}(t) \rangle = 0 \tag{33}$$

Low-pass $\times$ High-pass:

If $G(f) = 0$ for $|f| \geq W$, and $H(f) = 0$ for $|f| < W$, then:

$$\mathcal{H}[g(t)h(t)] = g(t)\hat{h}(t) \tag{34}$$

to compute the Hilbert transform of the product of a low-pass signal with a high-pass signal, only the high-pass signal needs to be transformed.

Amplitude-Modulated Signals:

$$\mathcal{H}[g(t)\cos(2\pi f_c t+\theta)] = \left[g(t) * \frac{\cos(2\pi f_c t)}{\pi t}\right]\cos(2\pi f_c t+\theta)+\left[g(t) * \frac{\sin(2\pi f_c t)}{\pi t}\right]\sin(2\pi f_c t+\theta). \tag{35}$$

Inverse Hilbert Transform: the inverse Hilbert transform is given by applying the Hilbert transform again, and negating the result

$$g(t) = -\mathcal{H}[\hat{g}(t)] = -\hat{g}(t)\frac{1}{\pi t}, \tag{36}$$

in general we need to include an additive constant to the result:

$$g(t) = -\mathcal{H}[\hat{g}(t)] = -\hat{g}(t)\frac{1}{\pi t} + c, \tag{37}$$

but assuming $g(t)$ and $\hat{g}(t)$ are zero mean signal the constant c is zero:

$$\mathcal{H}[\mathcal{H}[g(t)]] = -g(t) \tag{38}$$

Single-Sideband Modulation:

Define:

$$g_+(t) = \frac{1}{2}[g(t) + j\hat{g}(t)], \quad g_-(t) = \frac{1}{2}[g(t) - j\hat{g}(t)].$$

These give:

$$G_+(f) = G(f)u(f), \quad G_-(f) = G(f)u(-f).$$

Upper and lower sideband signals:

$$s_{USB}(t) = g(t)\cos(2\pi f_c t) - \hat{g}(t)\sin(2\pi f_c t),$$

$$s_{LSB}(t) = g(t)\cos(2\pi f_c t) + \hat{g}(t)\sin(2\pi f_c t).$$

Let's now discuss about some practical implementation in acoustic emission and impact localization of the Hilbert Transform.
The complex analytical signal is:

$$g_a(t) = g(t) + j \cdot \hat{g}(t) \tag{39}$$

This signal has only positive frequency components, and provides instantaneous amplitude and instantaneous phase.
The envelope is defined ad the module of the analytical signal:

$$E(t) = |g_a(t)| = \sqrt{g(t)^2 + \hat{g}(t)^2} \tag{40}$$

This curve follows the instantaneous energy content of the signal without alter the energy spectral density, highlighting the impulsive emission event and is particular robust to noise.
The envelope is particularly useful for:

- calculate the time of arrival thresholding directly on the amplitude thanks to the robustness to noise
- being always positive, the time interval between two zeros allows to highlight the temporal duration of the event
- distinguish multiple events and determine the time of arrival

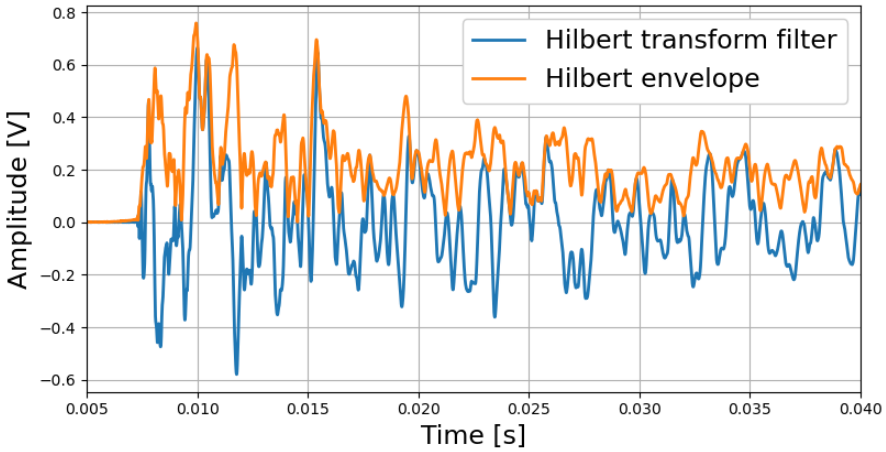


Figure 11: Comparison filtered signal and Hilbert envelope

3.2 Features Extraction

Once the signal has been filtered from the noise, theoretically the information remaining within it refers exclusively to the AE. In order to continue with the analysis it is necessary to extract that information from the signal in the form of characteristic quantities. Some of them related to AE have already been mentioned in Chapter-1, but in the following subsections we will focus on the TOA (defined in Section-1.1) extraction methods, subsequently implemented in the experimental analyses. In particular we will delve into the methodologies of threshold trespass, Maeda's Akaike information criteria (MAIC) and continuous wavelet transform.

3.2.1 Threshold Trespass

Currently, among the methods for picking the TOA of AE waveforms, the fixed-threshold method still holds a dominant position. This method determines the TOA as the moment when the signal amplitude first reaches a pre-set threshold. Due to its simple implementation process and relatively high picking efficiency, it is widely adopted in practical applications.

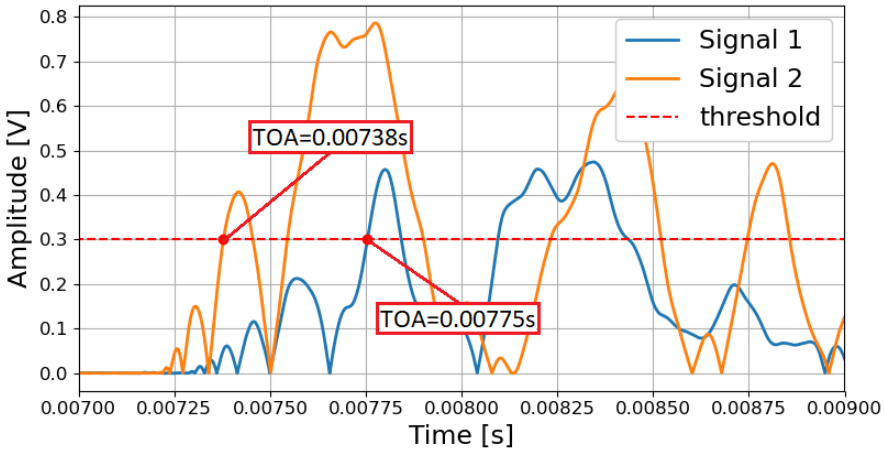


Figure 12: ToA identification using fixed-threshold method

However, for signals with a low-SNR, due to the influence of noise interference, the fixed-threshold method has obvious deficiencies in the accuracy of TOA picking [32], because using a high threshold level will lead to inaccurate time of arrival measurement, while a lower threshold value will increase the ability to pick the accurate waveform onset but also increases the risk of a false trigger.

3.2.2 Maeda's Akaike Information Criteria

The AIC function compares the signal entropy before and after each point in a signal and returns a minimum at the signal onset where the greatest difference is seen between the high entropy random noise seen prior to signal onset and the low entropy structured signal after onset. AIC is a robust measure to select the best approximating model of given models and is an Auto-Regressive (AR) time picking algorithm. This criterion has been widely used in seismology to investigate the onset

of seismic waves [33]. There are two different formulas when calculating AIC.
The first is:

$$AIC = -2 \cdot ln(L) + 2k \tag{41}$$

where L is the maximum likelihood estimate for the model and k is the number of fitted parameters in the model.
The other is:

$$AIC = n \cdot ln\left(\frac{RSS}{n}\right) + 2k \tag{42}$$

where the new parameter RSS is the residual sum of squares of the model and n is the sample size. Alternatively, the AIC can be calculated in a direct way of time series itself without using the AR coefficients, which was proposed by Maeda [31]. The minimum value of AIC represents the onset point.
The equation is defined as:

$$AIC(k) = k \cdot log(\sigma_1^2(x(1,k))) + (N-k-1) \cdot log(\sigma_2^2(x(k+1,N))) \tag{43}$$

where:

- $x = x_1, x_2, ..., x_N$: discrete amplitude values of recorded signal
- L: signal length (number of samples)
- N=L/2: half signal size
- $k = 1, 2, ..., N-1$: index of possible wave ToA
- first signal window variance: $\sigma_1^2(k) = \frac{1}{k}\sum_{i=1}^{k}(x_i - \mu_1(k))^2$
- second signal window variance: $\sigma_2^2(k) = \frac{1}{N-k}\sum_{i=k+1}^{N}(x_i - \mu_2(k))^2$
- mean value first signal window: $\mu_1(k) = \frac{1}{k}\sum_{i=1}^{k} x_i$
- mean value second signal window: $\mu_2(k) = \frac{1}{N-k}\sum_{i=1+k}^{N} x_i$

Basically the Maeda's AIC evaluate the difference in the variance between two signal window ($x(1 \to k)$ and $x(k \to N)$) varying the k parameter from 1 to N. The $AIC(k)$ function has a global minimum that represent the time value in which TOA occurs:

$$k_{TOA} = min(AIC(k)) \tag{44}$$

$$Time_{(TOA)} = t(k_{TOA}) \tag{45}$$

with t representing the time vector.

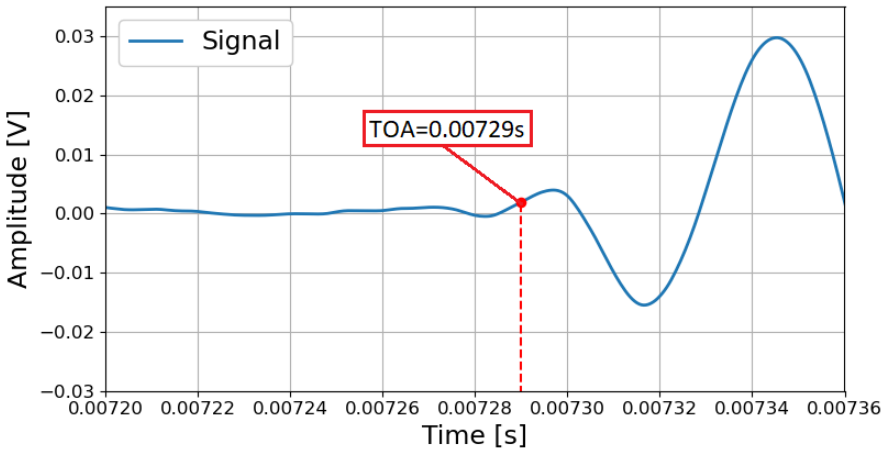


Figure 13: Maeda's AIC TOA detection

3.2.3 Continuous Wavelet Transform

The wavelet transform is a method to decompose a real signal f into a set of elementary waveforms that provide a way to analyze the signal by examining the wavelet coefficients. The continuous wavelet transform (CWT) of f is defined by:

$$CWT_f(a,b) = \int_{-\infty}^{\infty} f(x) \cdot \zeta_{a,b}^{-}(x)dx, \qquad a > 0,\ b \in R \tag{46}$$

where the mother wavelet is:

$$\zeta_{a,b}(x) = \frac{1}{\sqrt{a}} \cdot \zeta \cdot \left(\frac{x-b}{a}\right) \tag{47}$$

and ζ is the complex conjugate of $\zeta_{a,b}$, a is the scaling factor (dilation or compression), and b is the translation factor (time shift). The continuous wavelet transform has the property to be invertible if the condition of admissibility is respected:

$$C_\zeta = \int_{-\infty}^{\infty} \frac{|\zeta_{FT}|}{|v|} dv < \infty \tag{48}$$

with ζ_{FT} the Fourier transform of ζ.
Under this admissibility condition, the inverse wavelet transform can be calculated:

$$f(x) = \frac{1}{C_\zeta} \int_{a=0}^{\infty} \int_{b=-\infty}^{\infty} \frac{1}{|a|^2} \cdot CWT(a,b) \cdot \zeta_{a,b}(x) \quad da\ db \tag{49}$$

Examples of famous continuous wavelet[34]:

$$\zeta(x) = e^{-\pi x^2} \cdot e^{10i\pi x}$$

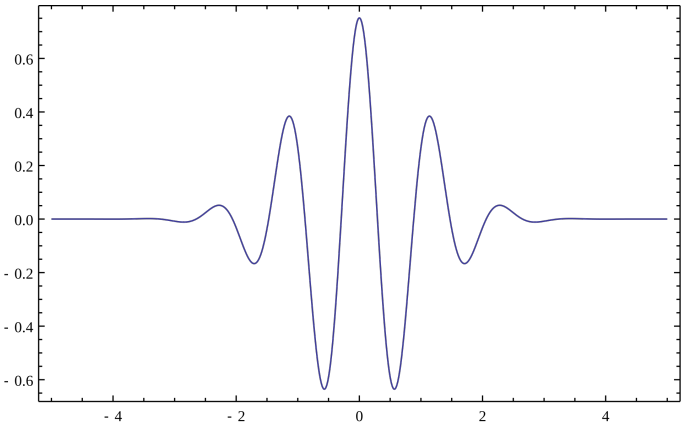


Figure 14: Morl mother wavelet

$$\zeta(x) = \frac{2}{\sqrt{3\sigma}\pi^{\frac{1}{4}}} \left(1 - \left(\frac{t}{\sigma}\right)^2\right) \cdot e^{-\frac{t^2}{2\sigma^2}}$$

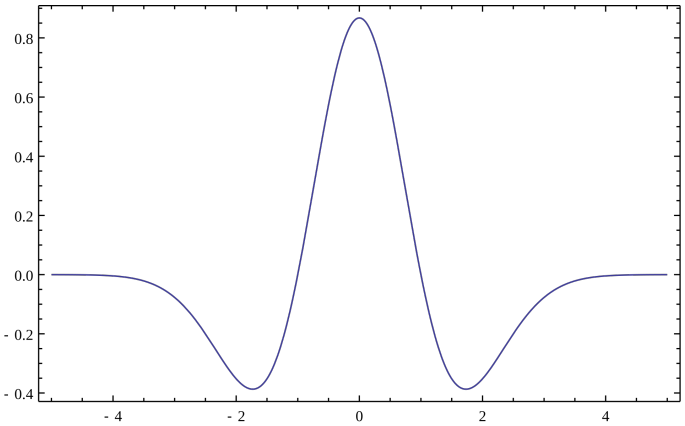


Figure 15: Ricker/Mexican-hat mother wavelet

Concerning the graphical representation of the wavelet coefficients of the CWT, is possible to plot the energy of the signal computed using the square modulus of the continuous wavelet transform, named the scalogram [35], which is equivalent to the spectrogram for wavelet.
The scale axis is expressed in frequency, using the relation between scale and frequency:

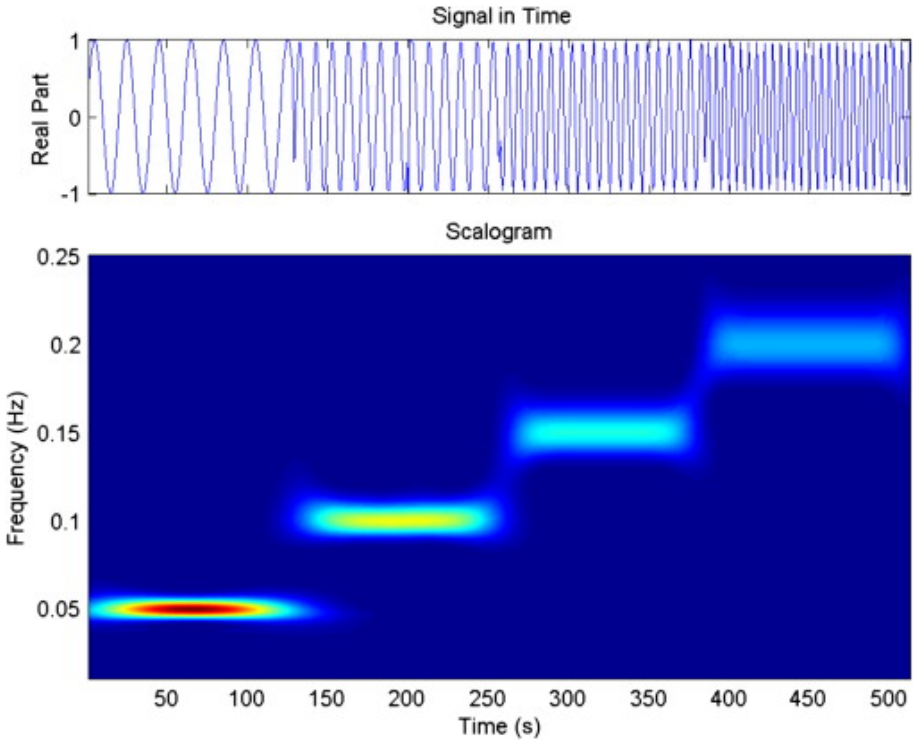


Figure 16: Scalogram (bottom) of the signal (top) composed of four sinusoidal functions at four different frequencies

Regarding the TOA extraction, we select a certain frequency window, the central frequency serves as the tunable parameter and the TOA correspond to the time when the CWT amplitude trespass an imposed threshold (another tuning parameter). This maximum CWT amplitude indicates that TOA is extracted from the dominant waves at that specific frequency and connects more closely to the wave group velocity[36].

4 Experimental Campaign

The experimental campaign took place entirely in the Braunschweig headquarters of the German national research center-DLR, within the lightweight systems laboratory, under the supervision of M.Sc. A. Bayoumi and M.Sc. M. Moix-Bonet.
During the campaign, two samples were examined, one is represented by an isotropic aluminum plate and the other is an anisotropic plate in composite material. Both plates had no structural complexity or noticeable defects.
The aim of the experiments was to obtain the signal data deriving from impact's AE on a grid of points reported on the plates, in order to constitute the database for the implementation of ΔT method (see Section-1.1) and for training of machine learning algorithms (see Section-1.2.
The aspects common to both experiments on the two samples concern the acquisition system, the sensors, instruments used and positioning of the plate.
Starting from the positioning of the plate, both were fixed at the four corners on an aluminum frame simulating the support boundary condition. The fixing of the plates took place in the work surface of a machine with a numerically controlled robotic arm, the position was not altered for the entire duration of the experiments.

Figure 17: Support frame and pins

The tool used to impact the surface is a spring-loaded plastic piston with a diameter of approximately 1 cm. it was chosen as a source for its consistency in impact energy and its ability not to damage the plates.

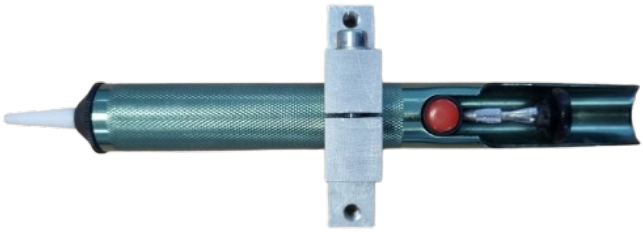


Figure 18: Plastic piston

The sensors used are receiving piezo-ceramic transducer in Lead zirconate titanate PIC255.
The acquisition system is National instrument® device with a maximum of 8 input channels. Before being recorded by the acquisition system and converted into digital, the signal coming from the sensors passes through a deamplifier with 25 to 1 ratio. The softwere used for data acquisition is `DAQ_RapidEF_20240605` property of DLR and developed by M.Sc. M. Moix-Bonet, the settings are:

— 0.13 *sec* acquisition window

— 2.0×10^6 sample per second

— 0.3 V threshold

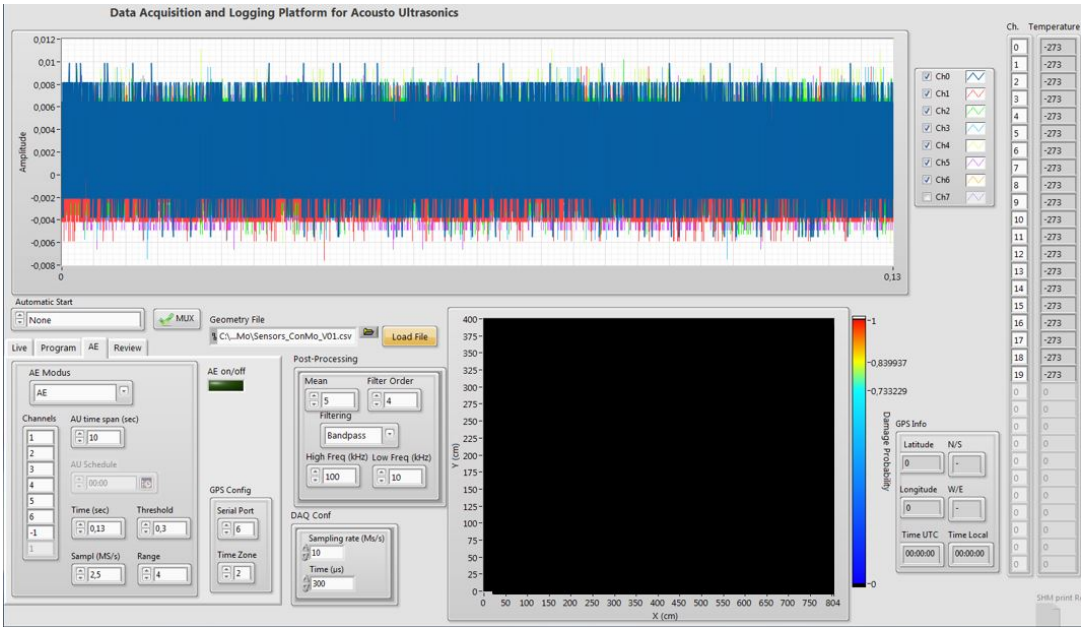


Figure 19: DAQ_RapidEF_20240605 layout

4.1 Aluminum Plate Experimental Set Up

The test article consist in a flat 1000×1000 mm aluminum plate without geometrical complexity or defect. The grid on which the impact tests were carried out is separated by 175 mm from the edges, with a total width of 650×650 mm.

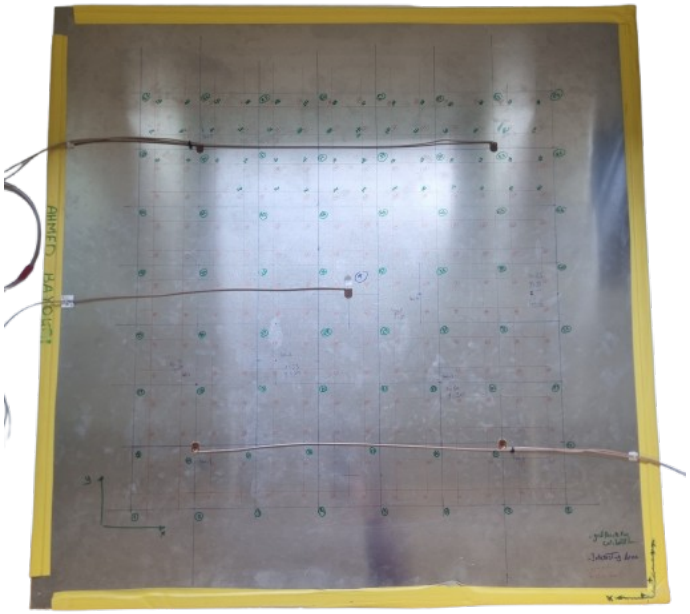


Figure 20: Aluminum plate

Referring to Figure-20, the bottom left corner of the grid represents the origin of the reference frame, the Y axis is upwards and the X axis is towards the right. The plate was discretized with 196 grid nodes each one separated by 50 mm increment on X and Y, 5 piezoelectric sensors are mounted at the following coordinates:

— Sensor 1: $x = 75$ mm, $y = 75$ mm

— Sensor 2: $x = 575$ mm, $y = 75$ mm

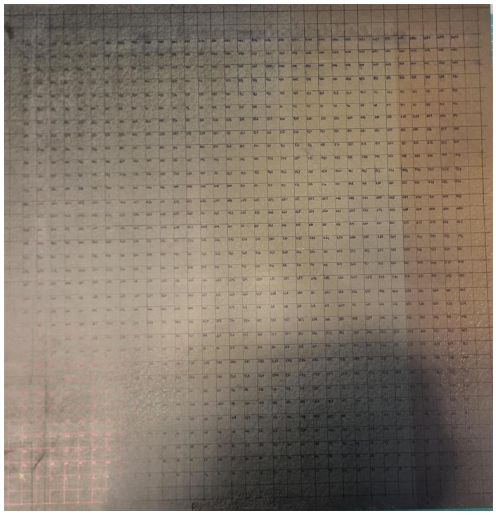
— Sensor 3: $x = 325$ mm, $y = 325$ mm

— Sensor 4: $x = 575$ mm, $y = 575$ mm

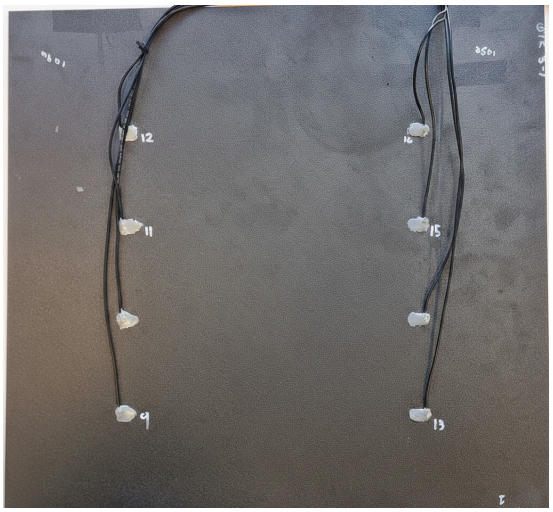
— Sensor 5: $x = 75$ mm, $y = 575$ mm

4.2 CFRP Plate Experimental Set Up

The CFRP plate was manufactured in German Aerospace Centre-DLR with a dimension of 0.57 m $\times$ 0.57 m and a nominal thickness of 2 mm. The corresponding ply thickness is 0.125 mm. The prepreg M21 / 34 % / UD134 / T700 / 300 from Hexply® was used to manufacture the plates with a quasi-isotropic (QI) layup of $[45\ /\ 0\ /-45\ /\ 90\ /-45\ /\ 0\ /\ 45\ /\ 90]_S$, material properties of a single unidirectional layer were measured based on standard test procedures [38].



CFRP plate I top



CFRP plate I bottom

Figure 21: CFRP plate I configuration

The grid is identified by the internal square of 480 mm $\times$ 480 mm with a distance of 45 mm between each side and the facing edge in order to reduce the boundary reflection. The grid node separation is 15 mm on X and Y for a total number of 1089 nodes.
The sensors used are 6 receiving piezo-ceramic, glued on the bottom of the plate in position:

- Sensor 1: $x = 80\ mm$, $y = 90\ mm$
- Sensor 2: $x = 80\ mm$, $y = 290\ mm$
- Sensor 3: $x = 80\ mm$, $y = 390\ mm$
- Sensor 4: $x = 400\ mm$, $y = 90\ mm$
- Sensor 5: $x = 400\ mm$, $y = 290\ mm$
- Sensor 6: $x = 400\ mm$, $y = 390\ mm$

Figure 22: Piston, plate and robotic arm

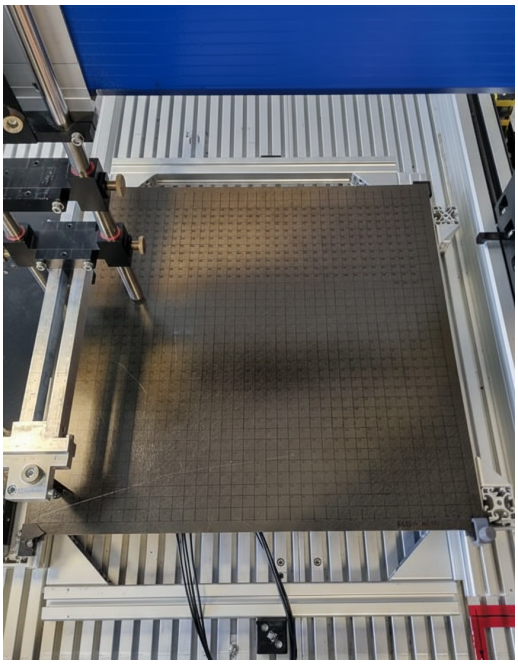


Figure 23: Complete CFRP plate experimental set up

5 Aluminum Plate Analysis

In this chapter we will discuss the methods of filtering the signal acquired through the experiments, than the implementation of the ΔT method and finally the validation of the method through testing.

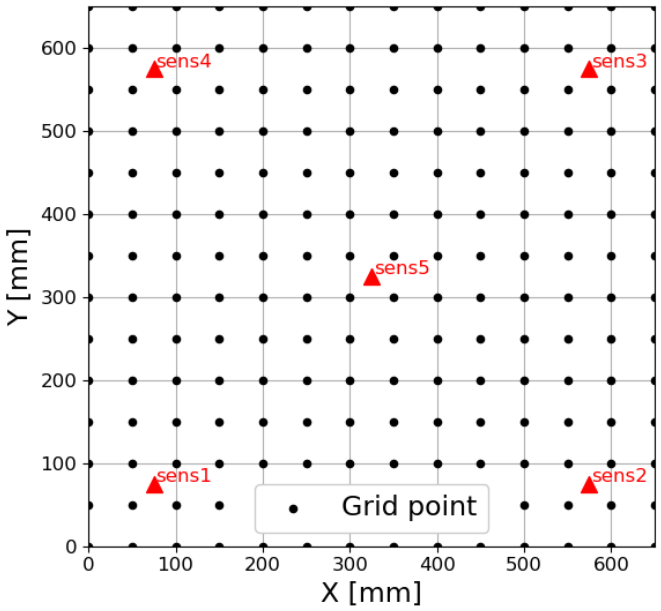


Figure 24: Aluminum plate digital model

The filtering sequence is obtained after investigation both on the type of filter and the sequence, the presented results permit to achieve the best accuracy so far.

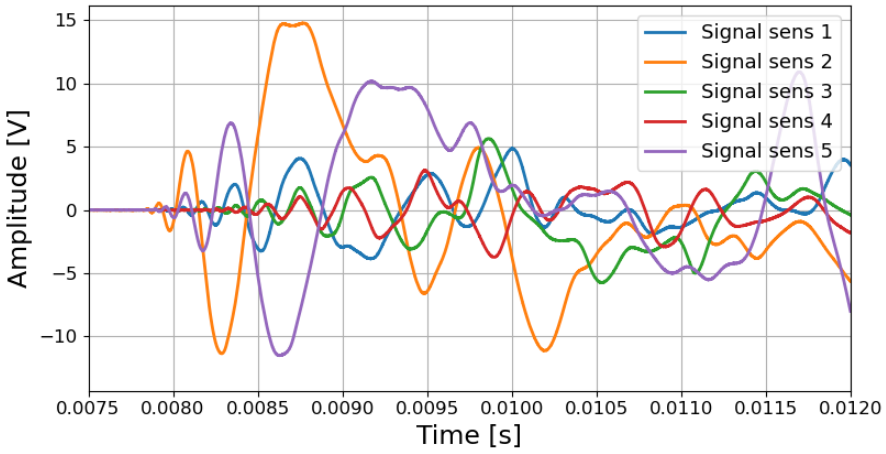


Figure 25: Unfiltered data acquired by 5 sensors on aluminum plate

The first filter is moving average on a 20 point window, symmetrical everywhere except for the initial and terminal part of the signal where the window become

asymmetrical due to lack of point in one of two sides.

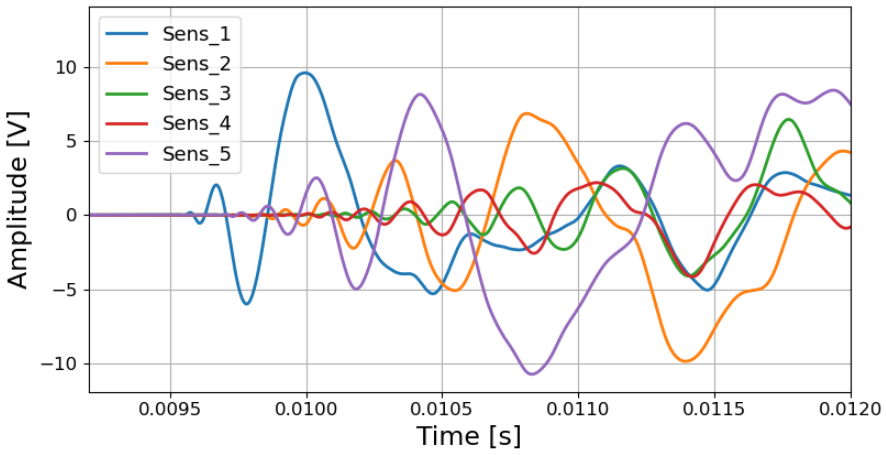


Figure 26: Moving average filter

The difference generated is not so relevant in general because these parts of the signal don't contain any sort of information about the impact event: in the initial part the plate is still waiting for the impact so the oscillations acquired are just noise with zero mean value, the terminal part represent the end of oscillations so again noise with zero mean value.

The next filter applied is Discrete Wavelet filter with sixth-order Daubechies wavelet (Db6).

Db6 coefficient	Value
h_0	0.332670552950
h_1	0.806891509311
h_2	0.459877502118
h_3	-0.135011020010
h_4	-0.085441273882
h_5	0.035226291882

With a sampling frequency of $f_S = 2 \times 10^6$ Hz we can scan the detail up to a various range of frequency:

Level j	Frequency Band $\left[\frac{f_s}{2^{j+1}}, \frac{f_s}{2^j}\right]$
d_1	[500.000, 1.000.000]
d_2	[250.000, 500.000]
d_3	[125.000, 250.000]
d_4	[62.500, 125.000]
d_5	[31.250, 62.500]
d_6	[15.625, 31.250]
a_6	[0, 15.625]

The filtering process is carry out by reducing the noise with a *soft threshold* based on standard deviation of the detail level where is buried most of the noise.
In this particular case the detail level is the 4-th so giving a fixed value of $1.2 \times \sigma(Detail\ level\ 4)$ as threshold we obtain a filter of this type (refer to Eq-21):

$$\text{soft thresholding}(x, \sigma(lv4)) = \begin{cases} \text{sign}(x) \cdot (|x| - \sigma(lv4)), & \text{if } |x| > \sigma(lv4) \\ 0, & \text{otherwise} \end{cases} \tag{50}$$

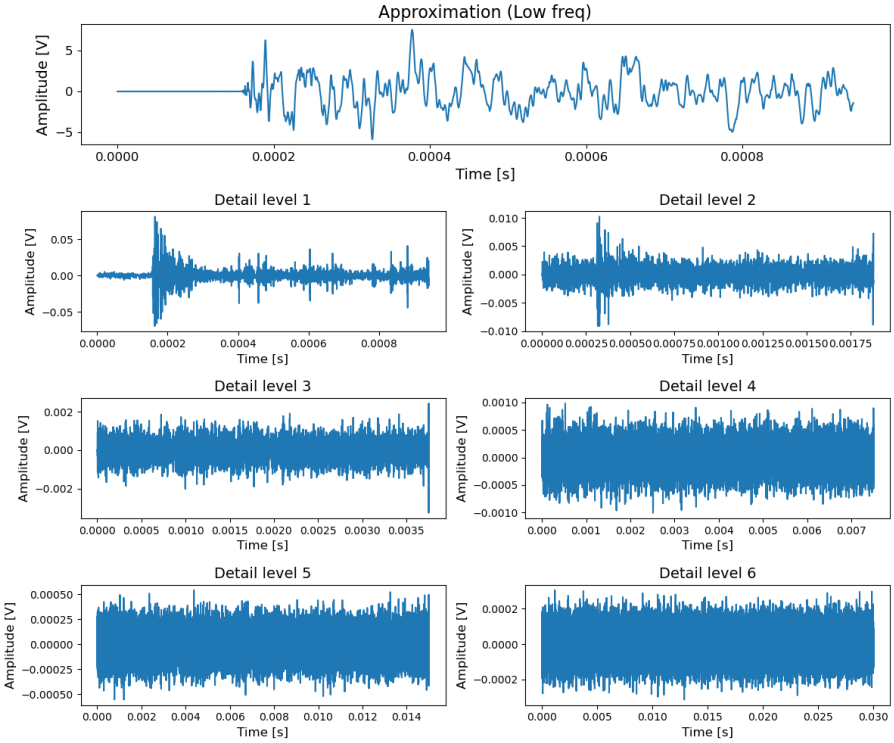


Figure 27: Approximation and detail Db6 filter

Suppressing all the noise contained in the detail level 4 before recomposition of the signal.

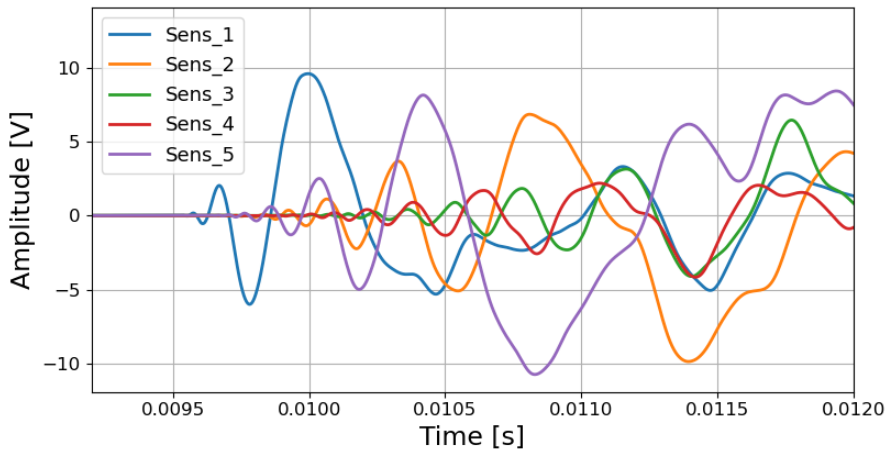


Figure 28: DWF with Db6 filter and soft thresholding

Last step before evaluating the time of arrival is applying the Hilbert transform filter.

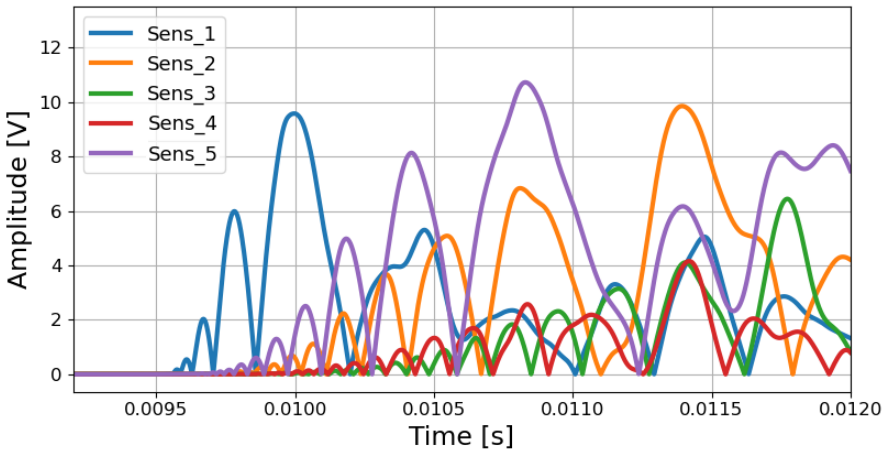


Figure 29: Envelope of Hilbert transform

The envelope obtained by the transform is particularly robust to noise so it can give information about the arrival in a more reliable way.

The methodology for AE localization is the ΔT method, fully described in Section-1.1. We now turn our attention to the TOA extraction method chosen for this case. The time of arrival computed in this section is based on trespassing of a fixed threshold related to the energy of the waves. The particular combination of filter is designed for maintaining the same amplitude of the peaks other than denoise the signal, in this way the time of arrival is unchanged.

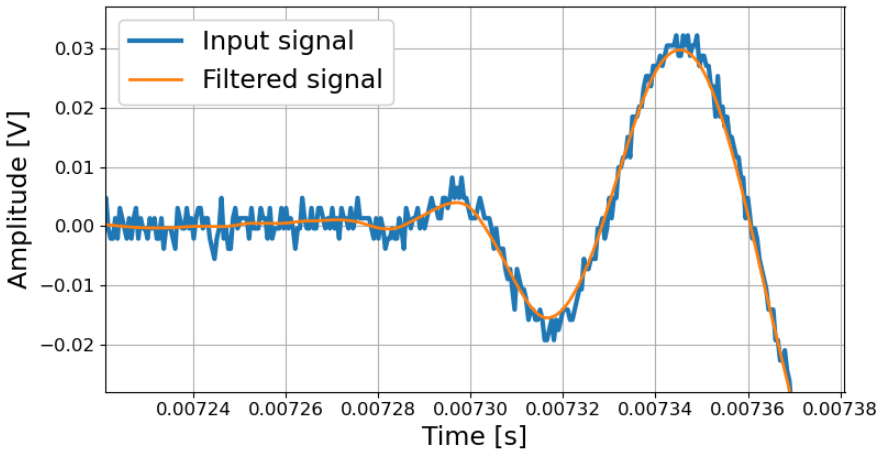


Figure 30: Comparison between unfiltered and filtered signal, focus on low reduction on amplitude peaks

The threshold that can better indicate the arrive of the impact wave for this test article is $0.6V$.

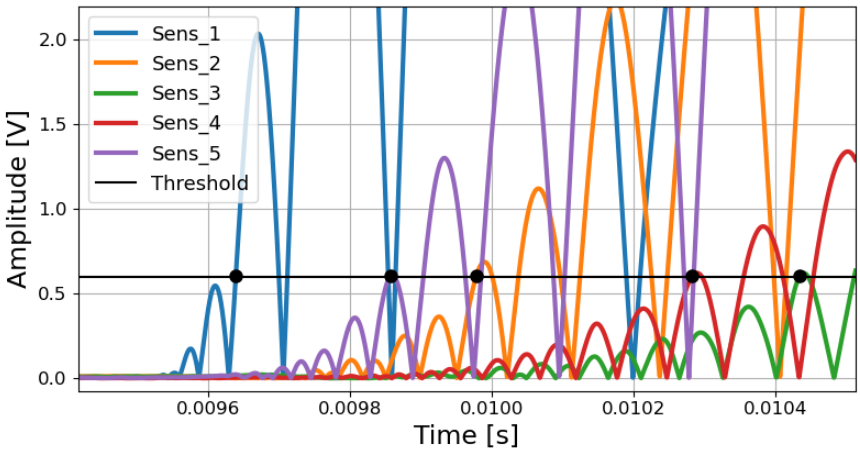


Figure 31: Hilbert envelope and threshold for calculating TOA

There is a redundancy of 4 impacts per each node so the final value of TOA associated to grid nodes will be the average value.
After obtaining the average arrival time of the wave at the various sensors for each grid point, we move on to calculating the differences (ΔT). Having 5 sensors available, the possible pairs are 10 (refer to Equation-2), therefore each node will have 10 ΔTs associated with it. After interpolating the data over 3000 point obtaining a finer mesh, we can isolate the iso-ΔT curves corresponding to the imposed sensor pair, a deviation (ϵ) of $\pm$ $10^{-8} sec$ on ΔT is set as tuning parameter (refer to Equation-3).

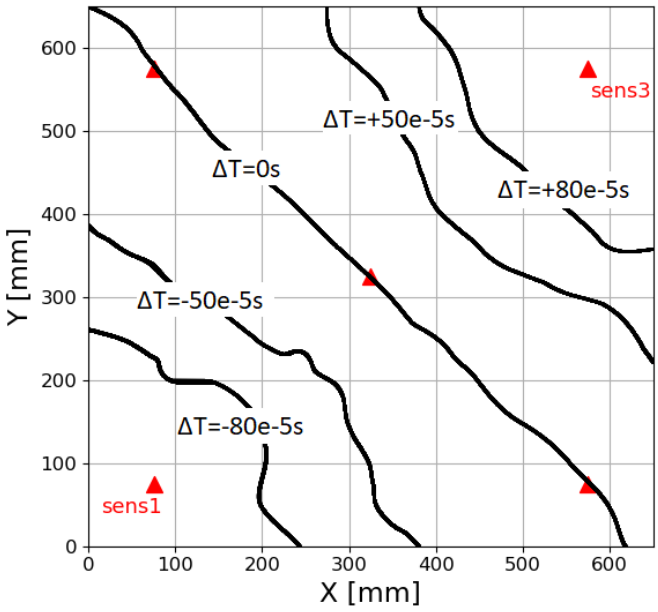


Figure 32: Iso-ΔT curves for sensors pair 1-3 associate to different ΔT

5.1 Testing

Proceed to test the developed algorithm with the following pipeline: (1) Test data ingestion, (2) test node TOA extraction, (3) iso-ΔT curves intersection and clustering for localization, (4) accuracy evaluation.

(1) Test data ingestion: a 10 point dataset is provided containing random grid nodes in order to test the accuracy in clustering and localization.

Table 3: **TEST POINTS**

ID	Coordinate [mm]
1	(150, 100)
2	(600, 150)
3	(150, 250)
4	(250, 300)
5	(250, 350)
6	(100, 400)
7	(500, 450)
8	(0, 550)
9	(550, 550)
10	(250, 650)

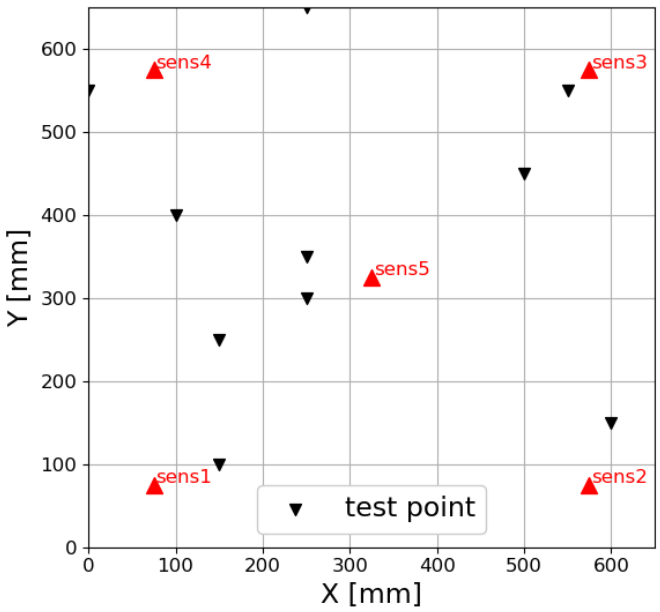


Figure 33: Test points on the plate configuration

(2) Test node TOA extraction: performed exactly as did with the grid point with same tuning parameters.

(3) iso-ΔT curves intersection and clustering for localization: selecting a test node

we can extract the relative TOA per sensor, then compute the ΔT per each sensor pair obtaining the respective 10 iso-ΔT curves.

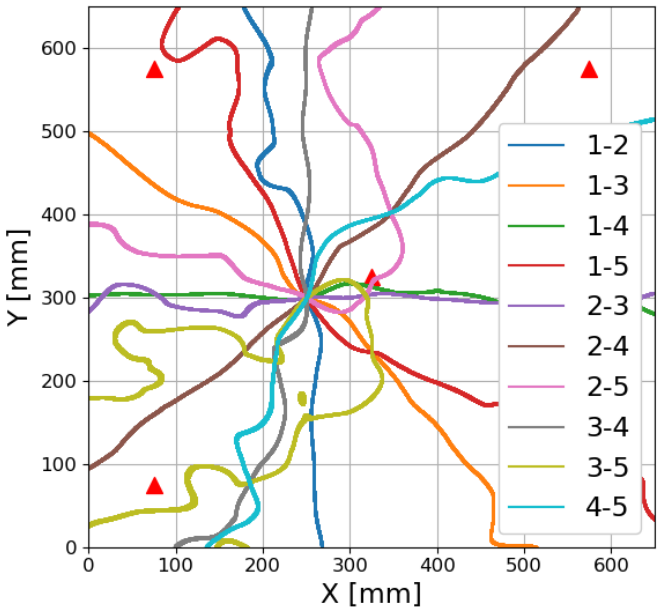


Figure 34: Iso-ΔT curves for single ΔT considering all sensors pairs

ID	Actual Coordinate [mm]	Predicted Coordinate [mm]
1	(150, 100)	(146.13, 102.75)
2	(600, 150)	(591.55, 149.11)
3	(150, 250)	(143.11, 248.38)
4	(250, 300)	(250.21, 299.69)
5	(250, 350)	(254.22, 355.41)
6	(100, 400)	(105.01, 400.76)
7	(500, 450)	(505.3, 448.46)
8	(0, 550)	(6.01, 548.03)
9	(550, 550)	(558.54, 548.85)
10	(250, 650)	(251.69, 647.65)

Table 4: Actual and predicted coordinate

Thanks to the isotropic property and absence of additional complexity or defects the curves generate only one common intersection that will be addressed as the localization point.
Repeating for all the test point we obtain:

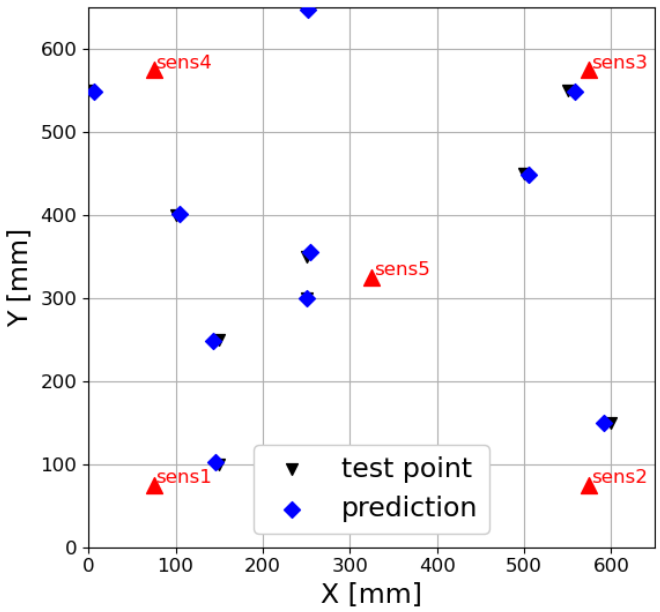


Figure 35: Test point location and predicted location comparison

(4) accuracy evaluation: defining the error as the euclidean distance between predicted location ad actual location,

$$error\ [mm] = \sqrt{(X_{prediction} - X_{test\ point})^2 + (Y_{prediction} - Y_{test\ point})^2} \tag{51}$$

and accuracy and precision as mean error and standard deviation respectively:

$$Accuracy\ (\mu) = \frac{\sum_{i=1}^{N} error_i}{N} \tag{52}$$

$$Precision\ (\sigma) = \sqrt{\frac{\sum_{i=1}^{N} (error_i - \mu)^2}{N}} \tag{53}$$

we can reassume the results:

ID	error [mm]
1	4.75
2	8.5
3	7.08
4	0.38
5	6.87
6	5.13
7	5.52
8	6.33
9	8.62
10	2.89
Accuracy	5.61 mm
Precision	2.39 mm

Table 5: Localization error per test point, accuracy and precision

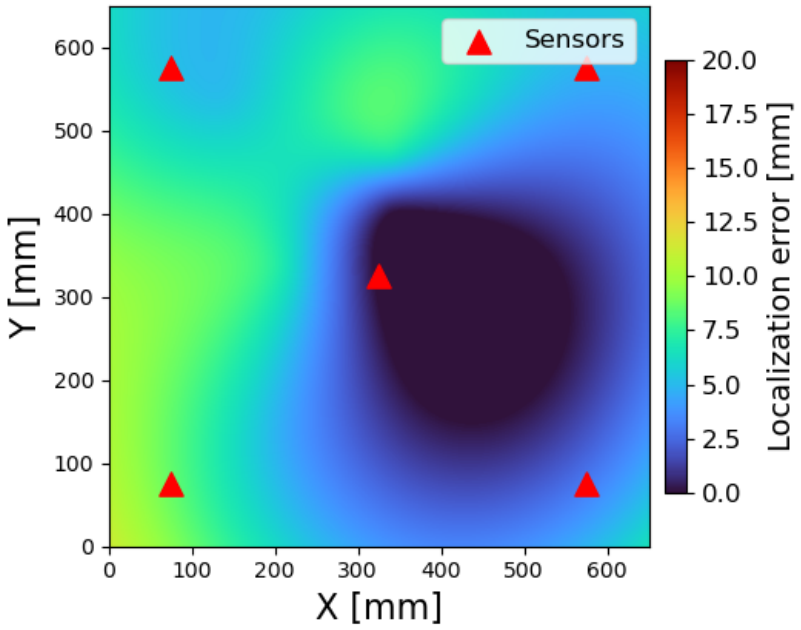


Figure 36: Interpolated localization error heat map for alluminum plate

5.2 Results Discussion

From a practical point of view we are interested in locating damage or defects starting from approximately 20 $mm \times 20$ mm dimension, and the results presented in the Table-5 prove a really high accuracy in localization so the objective can be considered satisfied. It is necessary to underline again the simplicity of the model (isotropic, no complexity or defect, high grid point number) so the carried out analysis is considered as a positive validation of the ΔT method algorithm implemented.
In the next analysis on a CFRP plate we expect the same method to result in a lower

accuracy due to the anisotropy of the material and the consequent difference in the wave propagation pattern. The objective will therefore be to refine the localization techniques in order to achieve an acceptable accuracy from an engineering point of view.

6 CFRP Plate Analysis

In this chapter, the challenge of localizing acoustic emissions on the CFRP plate presented in Section-4.2 will be addressed. During the discussion we will see first the implementation of the ΔT method, then the various techniques for enhancing the method which represent the novelty of this work. Subsequently, a random forest machine learning algorithm will be implemented and finally a comparison will be performed between the various techniques in terms of E_{9095}.

6.1 Application of ΔT Method

In this section is implemented the ΔT method for impact localization using all the possible sensors pairs on the CFRP plate; the process pipeline is presented step by step.

(1) Grid nodes signal processing: after acquired the experimental data through the acquisition system is required to trim the data in order to consider the first wave reaching the sensors and than filtering to increase the signal to noise ratio.
First step was led by Doc. Bayoumi using a DLR confidential code, while the filtering step is performed exactly as did in Section-5.

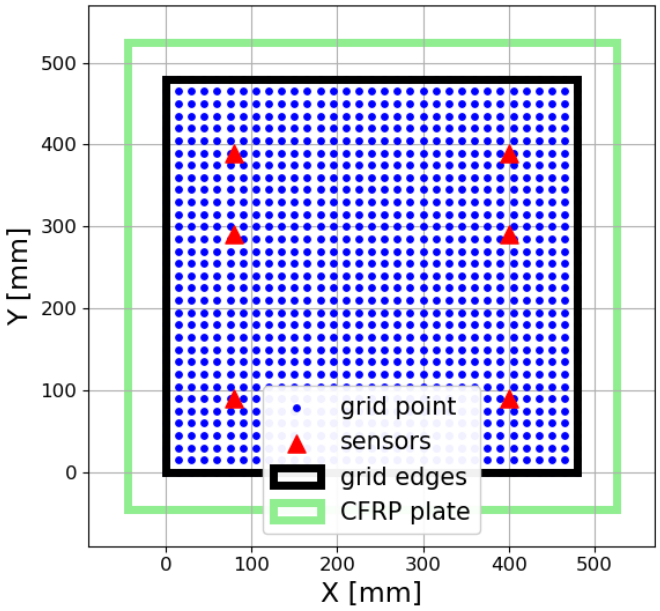


Figure 37: CFRP Plate I digital model

(2) TOA extraction: the TOA is computed using threshold trespassing on CWT of the signal (refer to Section-3.2.3), the implemented procedure is:

1. selecting the mother wavelet, in this case the *morl* shows better performance,
2. selecting the bandwidth where most of the energy is contained, for this signal the carrier frequency is around 1000 Hz so the band is selected as $[400, 1600]$ Hz,

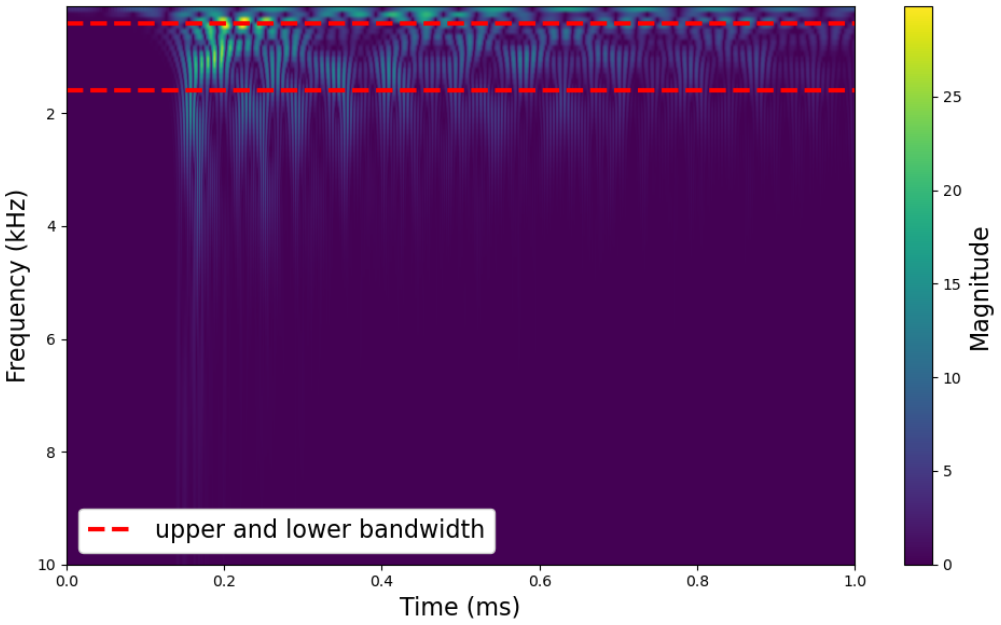


Figure 38: Spectrogram of impact signal

3. selecting threshold ratio, it is the tuning parameter for the TOA picking and represent the ratio between the amplitude threshold and absolute maximum signal amplitude inside the bandwidth, for this analysis is set at 0.1.

(3) ΔT computing: per each grid node there is a redundancy of 4 impacts, so the ΔT per sensors pairs is established as the mean value

$$\begin{aligned}
&\text{Impact 1)} \quad \Delta T^1_{m-n} = TOA^1_m - TOA^1_n \qquad m,n = sensors \quad ID \quad number \\
&\text{Impact 2)} \quad \Delta T^2_{m-n} = TOA^2_m - TOA^2_n \\
&\text{Impact 3)} \quad \Delta T^3_{m-n} = TOA^3_m - TOA^3_n \\
&\text{Impact 4)} \quad \Delta T^4_{m-n} = TOA^4_m - TOA^4_n
\end{aligned} \tag{54}$$

$$\text{Mean value:} \quad \bar{\Delta T}_{m-n} = \frac{\Delta T^1_{m-n} + \Delta T^2_{m-n} + \Delta T^3_{m-n} + \Delta T^4_{m-n}}{4}$$

(4) Create the interpolated ΔT map: starting from the ΔT computed over 1089 grid point, we refine the mesh interpolating the information on 2500 point using cubic spline.

(5) Test points data acquiring: test point dataset consists in 50 random nodes not belonging to the initial grid. Signal processing, TOA extracting and ΔT computing per sensors pairs is conducted as did for grid point in (1)-(2)-(3).

ID	Coord. [mm]	ID	Coord. [mm]	ID	Coord. [mm]	ID	Coord. [mm]
1	(233, 8)	14	(428, 428)	27	(113, 428)	40	(323, 248)
2	(67, 23)	15	(203, 443)	28	(323, 443)	41	(248, 323)
3	(157, 84)	16	(128, 293)	29	(473, 473)	42	(23, 383)
4	(413, 98)	17	(38, 443)	30	(473, 263)	43	(383, 38)
5	(293, 38)	18	(158, 203)	31	(338, 338)	44	(233, 203)
6	(38, 188)	19	(278, 383)	32	(173, 143)	45	(83, 323)
7	(278, 188)	20	(188, 353)	33	(383, 308)	46	(278, 473)
8	(458, 23)	21	(23, 278)	34	(383, 173)	47	(473, 143)
9	(68, 128)	22	(398, 263)	35	(188, 398)	48	(488, 398)
10	(338, 128)	23	(443, 338)	36	(398, 383)	49	(98, 83)
11	(458, 203)	24	(173, 38)	37	(263, 113)	50	(353, 413)
12	(83, 368)	25	(23, 68)	38	(113, 173)		
13	(293, 293)	26	(203, 263)	39	(98, 233)		

Table 6: Test points coordinates CFRP plate I

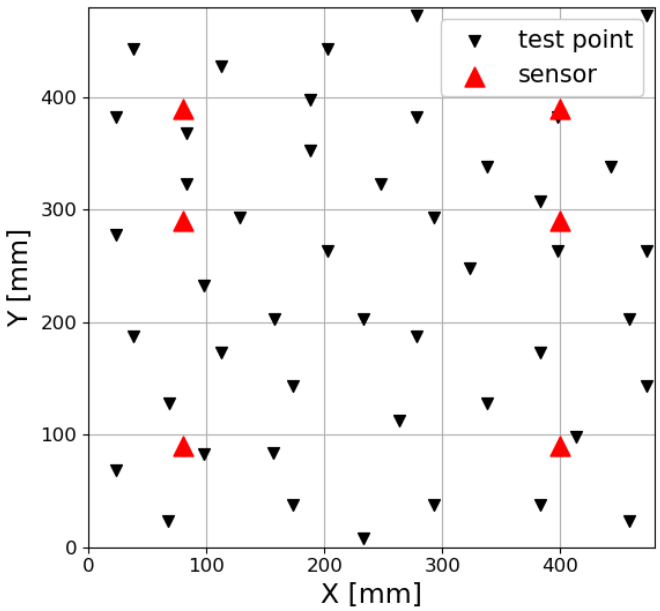


Figure 39: Test point and sensors configuration on CFRP plate I

(6) Nearest sensors identification: closer is the sensor to the AE source higher are the probability of acquiring a clean signal (less reflection, edge effects, structural defect), so in this analysis is given more importance to the information coming from sensors theoretically closer to the impact through the Nearest Neighbor logic.
Obviously the sensors are ordered from nearest to furthest ordering from lowest to highest associated TOA computed.
Example of sensors ordering for test point 35:

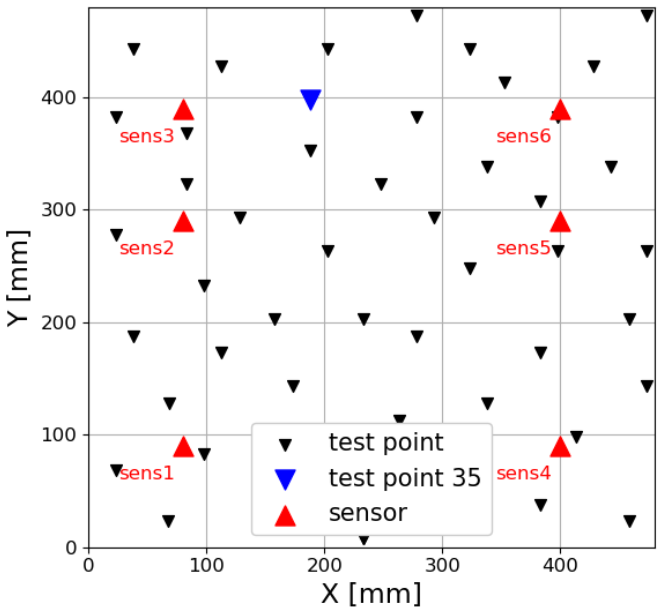


Figure 40: Test point 35 configuration

$$TOA_{\text{sens }1} = 0.0102 \text{ s}$$
$$TOA_{\text{sens }2} = 0.0089 \text{ s}$$
$$TOA_{\text{sens }3} = 0.0085 \text{ s}$$
$$TOA_{\text{sens }4} = 0.0103 \text{ s}$$
$$TOA_{\text{sens }5} = 0.0096 \text{ s}$$
$$TOA_{\text{sens }6} = 0.0093 \text{ s}$$

Ordered sensors sequence:

[Sens 3, Sens 2, Sens 6, Sens 5, Sens 1, Sens 4]

(7) Impact localization: per each test point are computed the 15 iso-ΔT curves related to all the possible combinations of sensors pairs. This time finding the curves intersection is not trivial due to the anisotropic characteristics of the plate, they are not intersecting around the same point as happened with the aluminum plate (Fig-34).

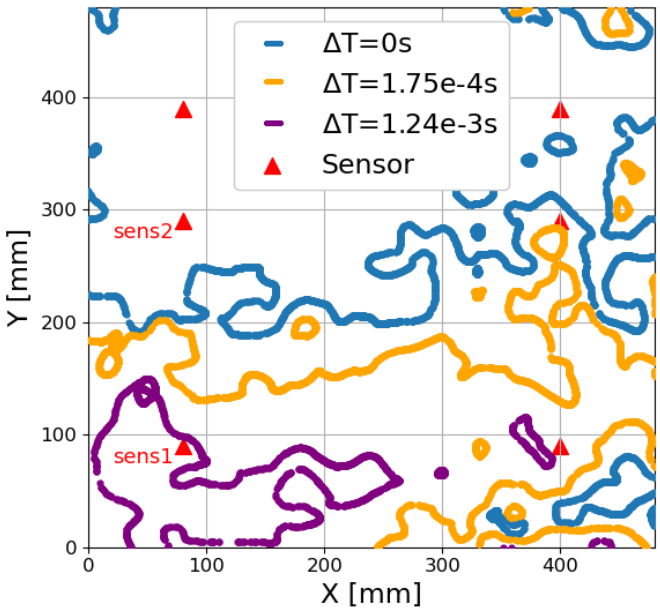


Figure 41: Iso-ΔT curves on CFRP plate considering sensors pair 1-2

As stated in point (6), so the ΔT curves are clearly distinct around the sensors, the further away we go the more they tend to lose coherence.
The proposed strategy to solve the problem is Nearest Neighbor logic on sensors, combined with geometric median clustering-based localization:

- **Geometric median clustering-based localization**: the iso-ΔT curves can intersect each other in more than one point, so the first step is evaluate all the intersections and group them in clusters using DBSCAN function, than choosing the more populated one and compute the iterative weighted centroid (see Subsection-9.1 in the appendix for further information). A fallback logic on curves can be implemented to manage possible unfeasible intersections.
- **Nearest Neighbor logic on sensors**: the information coming from sensors located theoretically closer to the source position is addressed as more valuable as explained in (6), so per each test point there will be 4 different estimation of X and Y obtained using first the nearest 3 sensors, than the nearest 4, than 5 and last one with 6. After acquired the 4 different location estimations the best accuracy so far is obtained using as final location guess the median value of the last 3 estimation related to the nearest 4, 5 and 6 sensors.

(8) Error evaluation: as stated in Section-5, the error is defined as the euclidean distance between actual source position and estimated one, while accuracy and precision are the mean error value and the standard deviation and $E_{90|95}$ is defined in Chapter-2:

- Accuracy = 41.47 mm
- Precision = 40.50 mm
- $E_{90|95}$ = 81.27 mm

ID	Error [mm]	ID	Error [mm]	ID	Error [mm]	ID	Error [mm]
1	115.77	14	48.05	27	58.45	40	32.49
2	56.17	15	23.26	28	35.88	41	56.96
3	11.19	16	23.76	29	214.35	42	12.53
4	21.10	17	36.68	30	79.24	43	10.19
5	91.256	18	26.05	31	42.81	44	9.14
6	5.88	19	16.42	32	13.03	45	14.24
7	46.10	20	9.10	33	38.15	46	61.98
8	125.21	21	34.15	34	12.88	47	102.07
9	9.06	22	5.42	35	19.89	48	125.38
10	10.66	23	59.43	36	7.43	49	24.82
11	20.87	24	21.58	37	42.29	50	51.18
12	91.78	25	58.07	38	12.35		
13	7.28	26	8.01	39	13.42		

Table 7: Error on test point localization for CFRP plate I

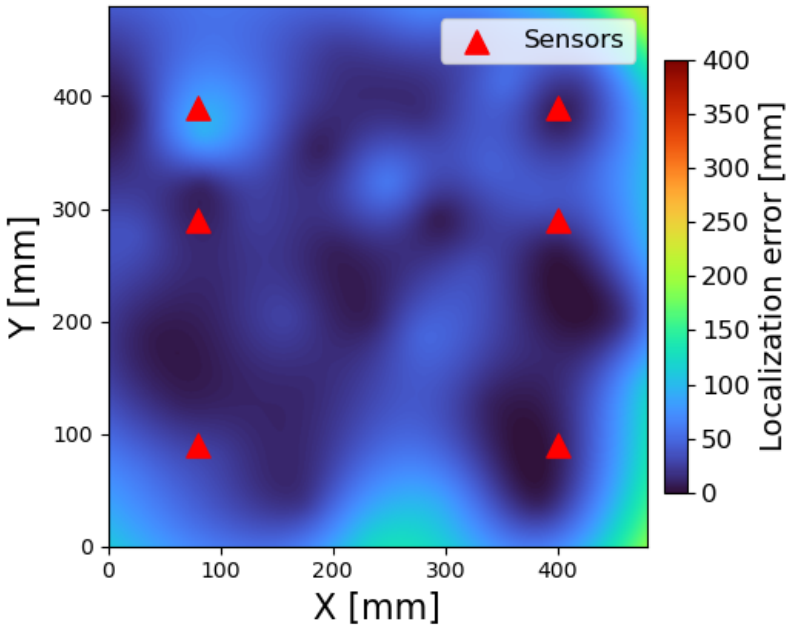


Figure 42: Interpolated localization error heat map

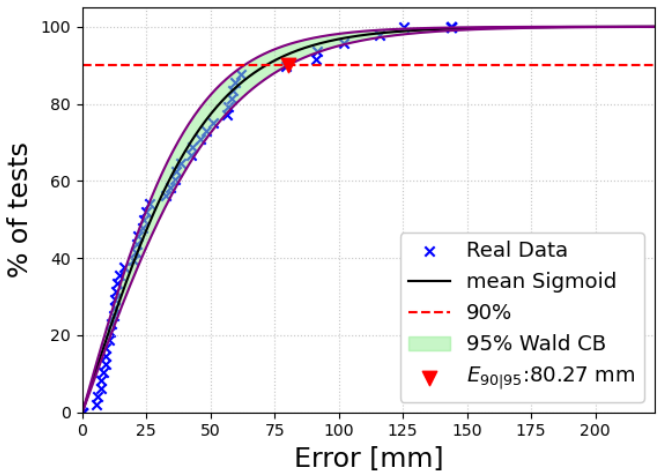


Figure 43: Regression sigmoid with 95% confidence interval and $E_{90|95}$

Focusing on localization performance in the area enclosed by the sensors: $80 \leq X \leq 400$ ad $90 \leq Y \leq 390$:

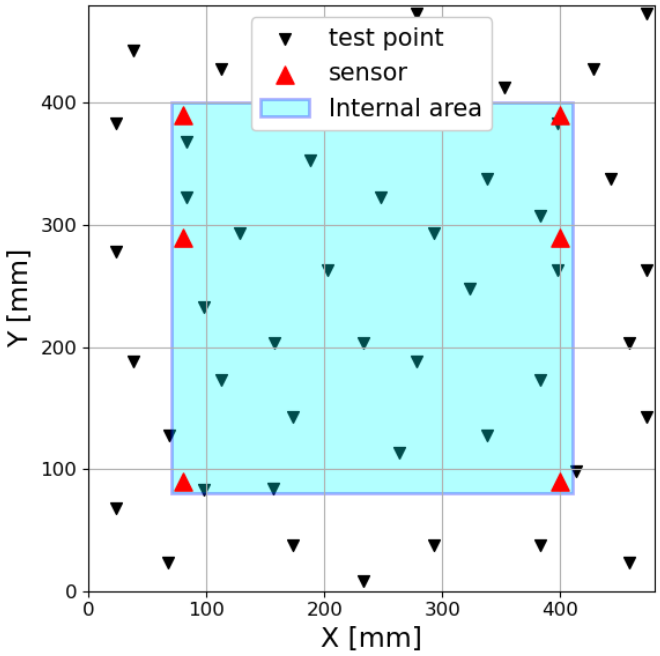


Figure 44: Internal area configuration

- Accuracy = 24.53 mm
- Precision = 20.83 mm
- $E_{90|95} = 45.08\ mm$

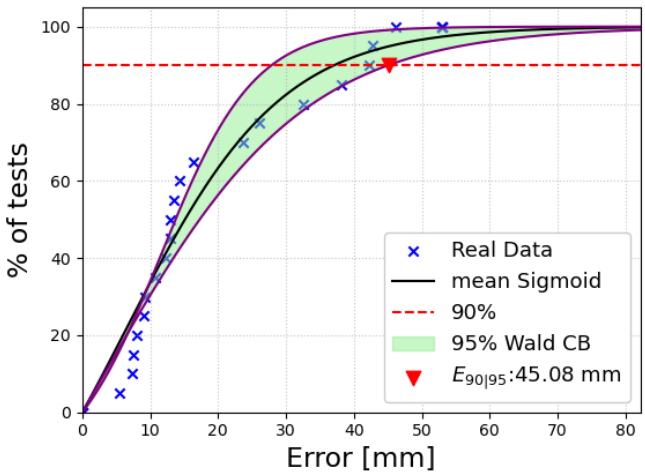


Figure 45: regression sigmoid with 95% confidence interval and $E_{90|95}$ referred to internal area localization

6.2 Sensors Pairs Region of Influence

In this section we will discuss the concept of region of influence, birth of the idea and practical implementation. As explained in Subsection-6.1 the information related to the nearest sensors is considered more valuable, based on this consideration seems reasonable to completely exclude the furthest sensors, and relative possible sensors pairs, from the localization process.
Sensors pair region of influence is defined as the portion of plate where the iso-ΔT curve relate to a specific sensor pair is proven be necessary to better localize the AE source.

In order to identify this regions we should rely on a broader test dataset, so each grid point in turn will be converted as test point and then localized according with the following pipeline:

1. Remove a grid node, one at a time, from the initial mesh, and consider it as a test point
2. Apply the ΔT method in order to localize the removed grid point as described in Subsection-6.1, but, instead of extract only one estimated location, we will save the results obtained with all the possible pair related to the nearest 3, 4, 5 or 6 sensors,
3. Evaluate the sensors pairs needed to achieve the lowest localization error,
4. Iterate the process for each grid node obtaining 15 map that shows where each one of the possible sensor pair is most effective in localization.

In Figure-46 each red dot represent a test point which the best accuracy in localization is achieved using all the 15 possible sensors pairs, the black dot instead indicate a test point which is better localized with less than 15 sensors pairs.

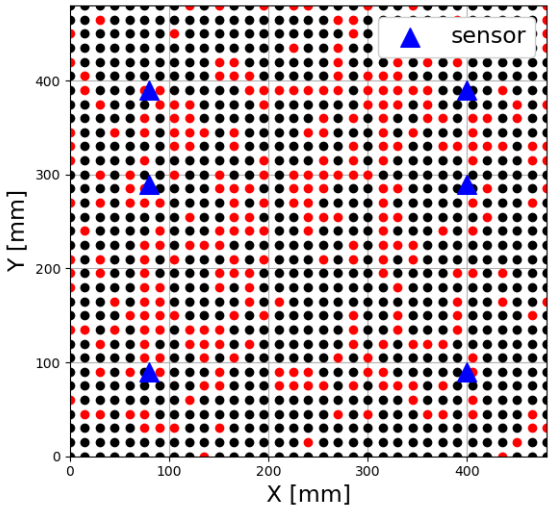


Figure 46: Grid point localization scatter plot

The red dot are 26.6 % of the total, supporting the idea of sensors region of influence existence.

We can now proceed to report per each sensor pair the same type of scatter plot but now the red dot identify a test point where for reaching the best accuracy the certain pair is necessary.
In this section is reported only the sensor pair 1-2 as example, the remaining others can be found in Subsection-9.2 in the appendix.

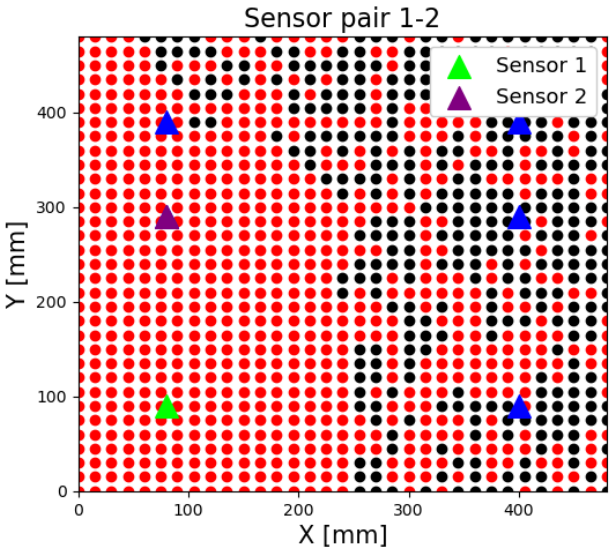


Figure 47: Sensor pair 1-2 scatter plot

As we could expect is formed a clearly distinguishable region of clustered red dots where the sensor pair 1-2 is most effective in localization. In this analysis the region of influence is constrained to embed at least 70 % of red dot, this choice is the result of a tuning process, for a more robust implementation is suggested the use of ANNs or genetic algorithms.

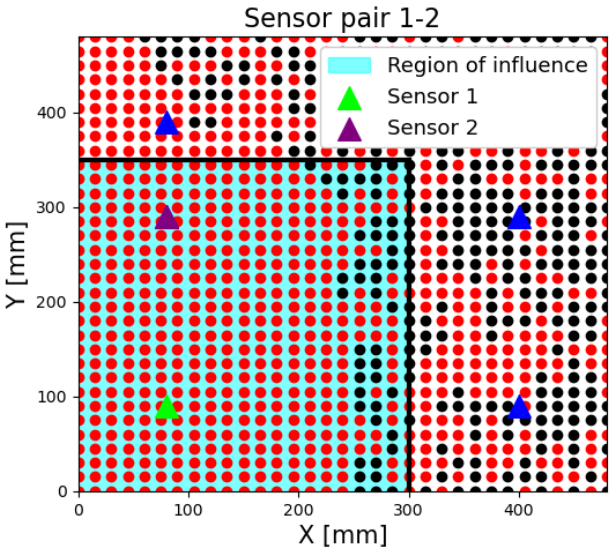


Figure 48: Sensor pair 1-2 region of influence

To include the sensors region of influence logic in the ΔT method algorithm each iso-ΔT curve related to a specific sensor pair will be trimmed outside the corresponding region.

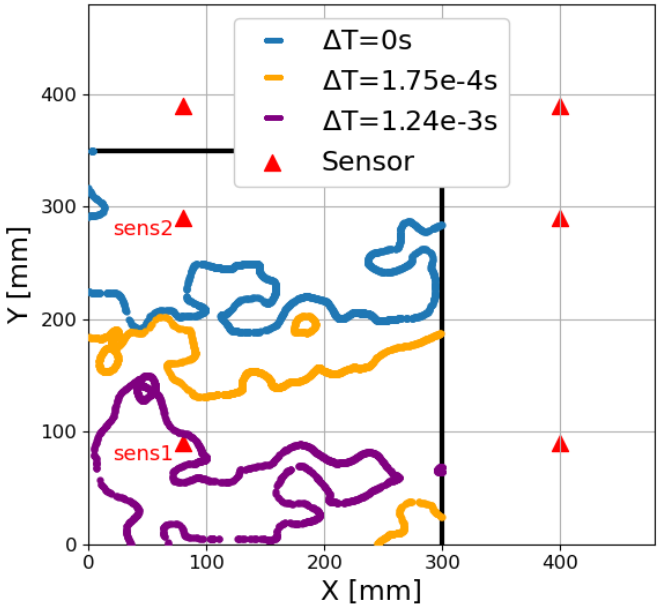


Figure 49: Iso-ΔT curves on CFRP plate considering sensors pair 1-2 inside the region of influence

As can be seen from Fig-49, now the remaining parts of the curves have a much more stable and defined trend.

Proceeding to test we will focus on improving the algorithm performance inside the area enclosed by the sensors.

Overall error assessment:

- Accuracy: 64.25 mm
- Precision: 57.61 mm
- $E_{90|95} = 116.84$ mm

ID	Error [mm]	ID	Error [mm]	ID	Error [mm]	ID	Error [mm]
1	141.61	14	53.28	27	70.85	40	22.77
2	178.64	15	27.24	28	33.63	41	12.51
3	7.30	16	14.73	29	208.43	42	167.88
4	17.58	17	153.23	30	104.96	43	66.82
5	59.65	18	18.95	31	14.12	44	7.97
6	127.84	19	10.31	32	10.22	45	5.35
7	10.92	20	9.67	33	17.35	46	85.94
8	69.40	21	156.79	34	17.58	47	35.87
9	75.53	22	14.18	35	9.819	48	25.89
10	2.31	23	78.21	36	8.83	49	73.33
11	120.46	24	16.65	37	78.17	50	46.90
12	18.22	25	310.49	38	9.10		
13	11.06	26	23.73	39	18.46		

Table 8: Error on test point localization for CFRP plate I using sensors region of influence

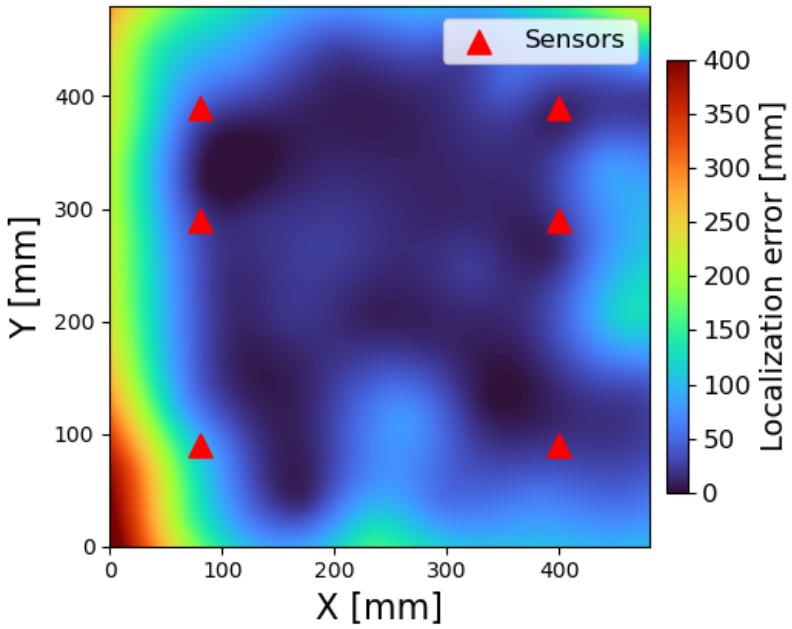


Figure 50: Interpolated localization error heat map with sensors region of influence

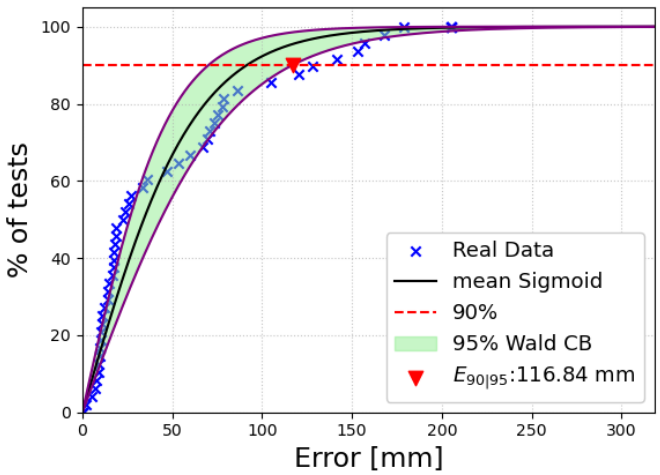


Figure 51: Regression sigmoid with 95% confidence interval and $E_{90|95}$ referred to whole plate

Error assessment inside internal area:

- Accuracy: 16.20 mm
- Precision: 14.51 mm
- $E_{90|95} = 20.07$ mm

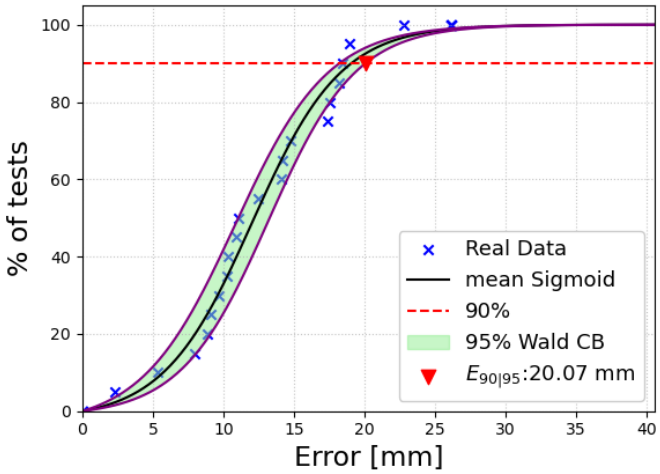


Figure 52: Regression sigmoid with 95% confidence interval and $E_{90|95}$ referred to localization with region of influence inside internal area

6.3 Combining Region of Influence and Classical ΔT Method

Looking at the results in terms of accuracy and precision obtained in Section-6.1 and Section-6.2, we can notice that the firs ΔT method algorithm implemented shows less error and more homogeneity in general allover the plate instead the second algorithm show an enhanced localization capability only inside the the area enclosed

by sensors. The logical continuation is to find a way of combine both approach in order to achieve the performance of region of influence inside the area enclosed by sensors and the performance of standard ΔT method outside.

The solution proposed is to localize with sensors region of influence the points that are estimated to be within sensors area, the one outside will be localized with the ΔT method algorithm presented in Section-6.1. It is therefore necessary a **double step localization**:

1. First step: we use the standard ΔT method (Section-6.1) algorithm for localizing the AE source. This choice is made based on the less mean error and more homogeneity of the error allover the plate.
2. Second step: if the estimated AE position is inside the sensors area, defined as internal area: $70 \leq X \leq 410$ and $80 \leq Y \leq 400$ (10 mm more on each side as tolerance respect to Section-6.1), the source will be re-localized with the sensors region of influence algorithm, otherwise the first localization is considered definitive.

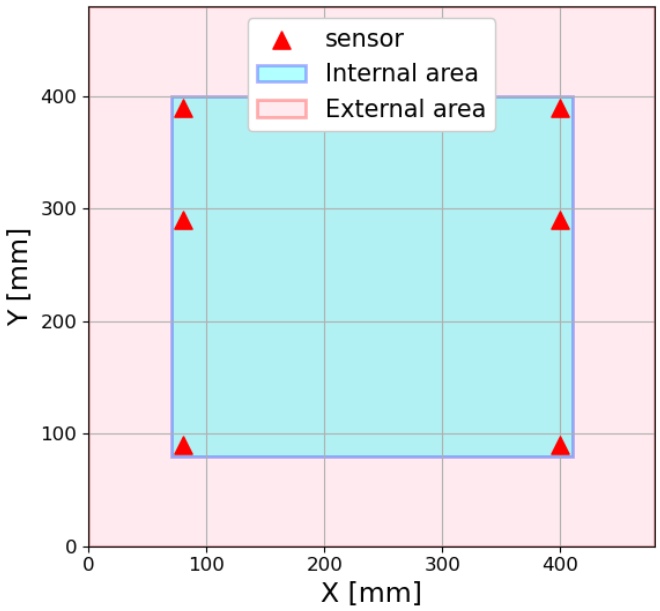


Figure 53: Internal and external area division

Using this combination the results obtained over the whole plate are:

- Accuracy: 38.00 mm
- Precision: 39.74 mm
- $E_{90|95} = 70.70$ mm

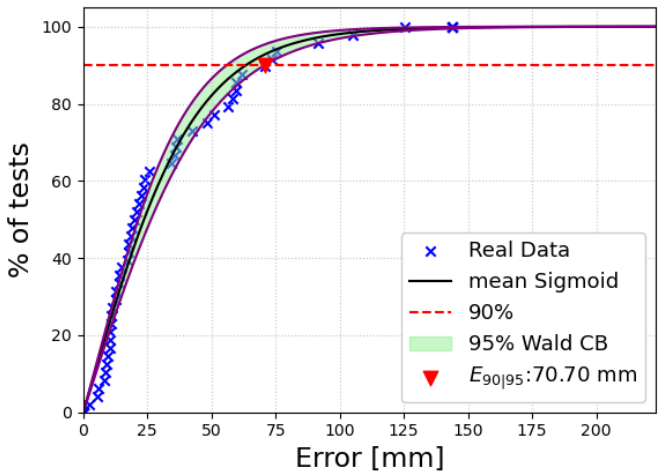


Figure 54: Regression sigmoid with 95% confidence interval and $E_{90|95}$ obtained using combined logic referred to whole plate

in particular in the internal area:

- Accuracy: 17.92 mm
- Precision: 17.99 mm
- $E_{90|95} = 20.83$ mm

Therefore, a general improvement in localization performance can be noticed.

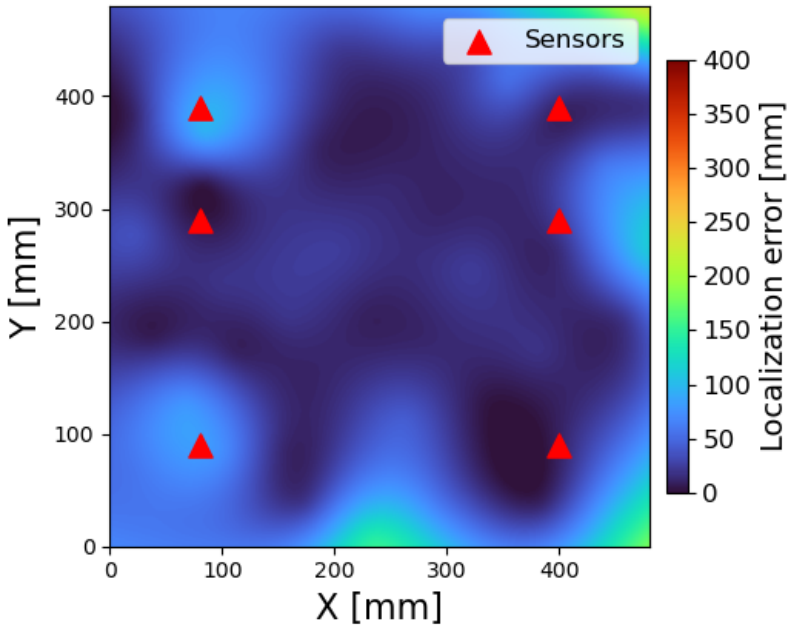


Figure 55: Interpolated localization error heat map with combined logic

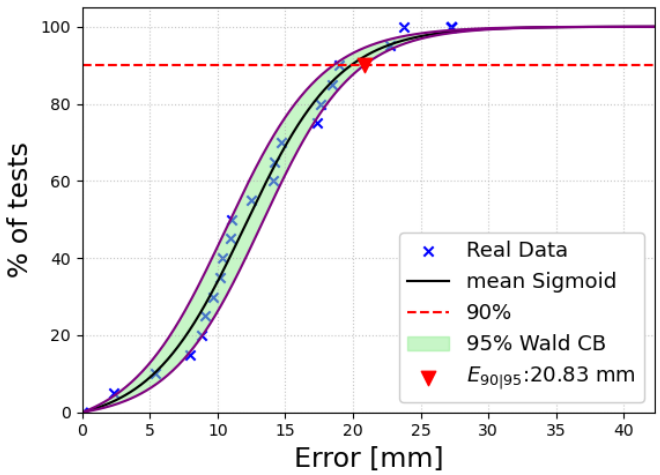


Figure 56: Regression sigmoid with 95% confidence interval and $E_{90|95}$ obtained using combined logic referred to internal are

6.4 Grid point reduction

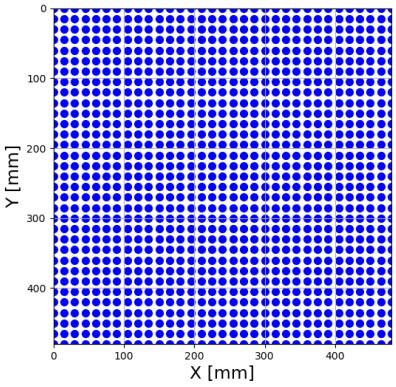
Next step to validate the methodology developed is reduce the grid point number and observe the scaling in accuracy, precision and POL.
In this section 2 analysis are proposed:

1. Reducing the points per row, widening the spacing on X and leaving unchanged the spacing on Y direction,
2. Changing the spacing among points both on X and Y removing rows and columns from the grid.

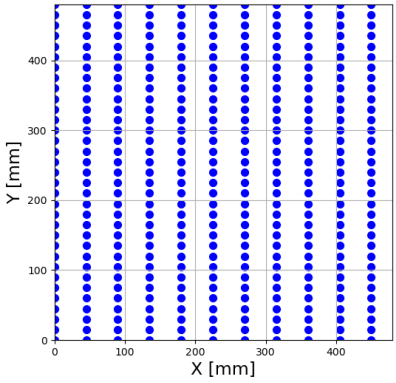
The following results are presented in terms of E_{9095}, for a complete overview of the accuracy and precision indicators please refer to Section-9.3 in the appendix.

6.4.1 Reducing points per row

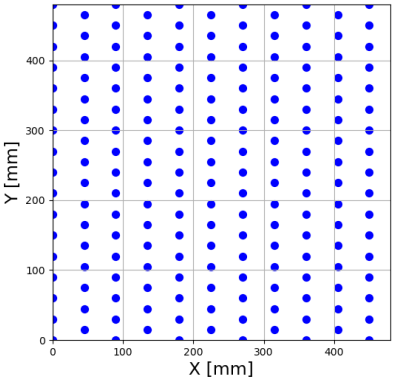
The algorithm's robustness to row grid point reduction is tested starting considering first 1 point each 2 per row (545 nodes, 50% of total) up to 1 each 6 (182 nodes, 17% of total).



(a) Initial grid configuration



(b) 1 to 3 reduction rate



(c) 1 to 6 reduction rate

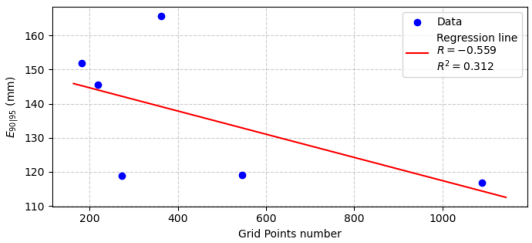
Figure 57: Examples of grid node reduction (1)

In the following results table the symbols refer to:

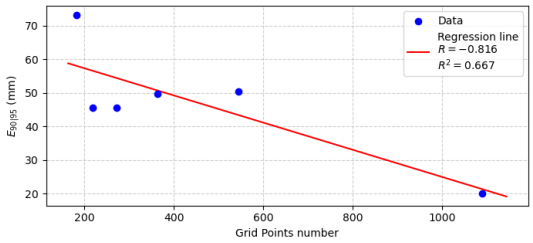
- $E_{90|95\ RWP}$: radius of the area associated with 95% confidence of localization considering the region of influence approach over the whole plate
- $E_{90|95\ RIA}$: : radius of the area associated with 95% confidence of localization considering the region of influence approach over the internal area
- $E_{90|95\ NRWP}$: : radius of the area associated with 95% confidence of localization considering the standard approach over the whole plate
- $E_{90|95\ NRIA}$: : radius of the area associated with 95% confidence of localization considering the standard approach over the internal area
- $E_{90|95\ CWP}$: : radius of the area associated with 95% confidence of localization considering combined logic over the whole plate
- $E_{90|95\ CIA}$: : radius of the area associated with 95% confidence of localization considering combined logic over the internal area

Radius [mm]	1089 points	545 points	363 points	273 points	218 points	182 points
$E_{90\|95\,RWP}$	116.84	119.16	165.69	118.88	145.49	151.79
$E_{90\|95\,RIA}$	20.07	50.34	49.76	45.56	45.63	73.15
$E_{90\|95\,NRWP}$	80.27	90.02	129.80	109.77	111.05	127.68
$E_{90\|95\,NRIA}$	45.08	53.14	54.09	60.78	44.97	67.10
$E_{90\|95\,CWP}$	70.70	94.51	144.59	102.79	118.84	136.46
$E_{90\|95\,CIA}$	20.83	58.94	63.28	68.99	58.61	81.57

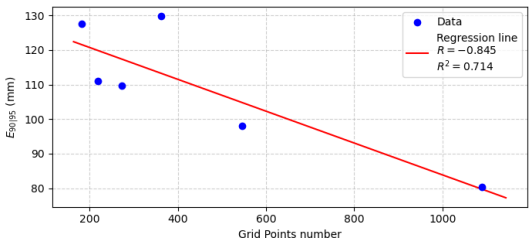
Table 9: Grid point reduction results (1)



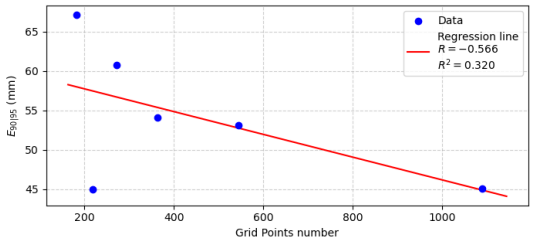
$E_{90|95\,RWP}$ regression



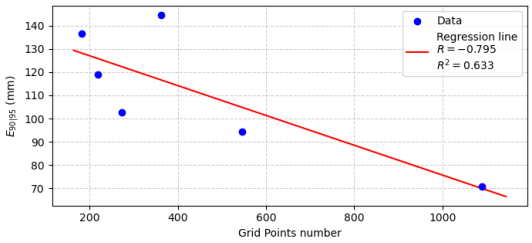
$E_{90|95\,RIA}$ regression



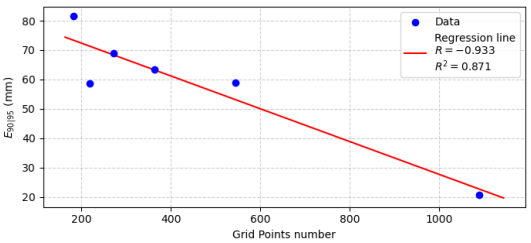
$E_{90|95\,NRWP}$ regression



$E_{90|95\,NRIA}$ regression



$E_{90|95\,CWP}$ regression

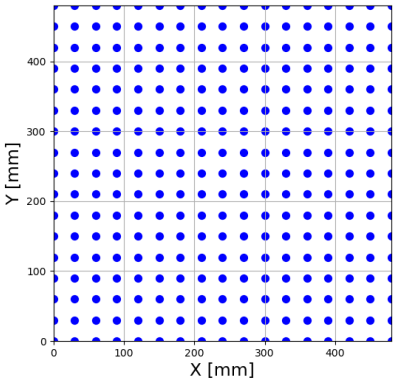


$E_{90|95\,CIA}$ regression

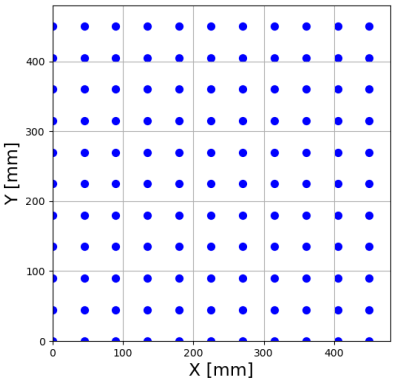
As shown in figures above the increase in $E_{90|95}$ is linear with grid points reduction (not exponential as could be expected), and the slope is less than the bisector of second quadrant so this results prove the robustness of the algorithm in pre-established conditions.

6.4.2 Reducing rows and columns

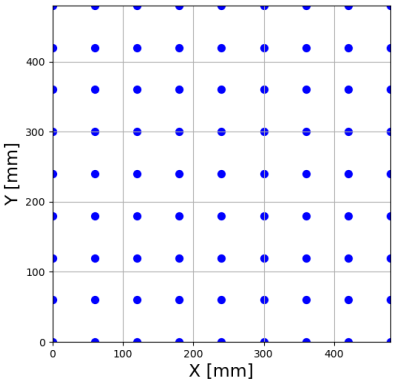
Proceeding now to test the robustness to rows and columns reduction starting with removing 2 rows and 2 columns each 3 (289 nodes, 26.5% of total), than 3 each 4 (121 nodes, 11% of total) and finally 4 each 5 (81 nodes, 7.4% of total).



(a) $\frac{2}{3}$ row and column removed



(b) $\frac{3}{4}$ rowand column removed



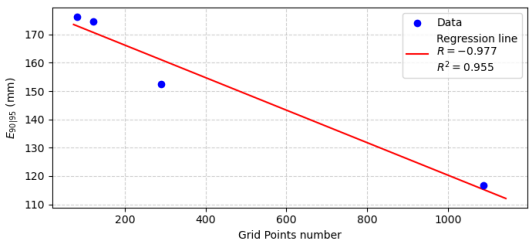
(c) $\frac{4}{5}$ row and column removed

Figure 61: Examples of grid node reduction (2)

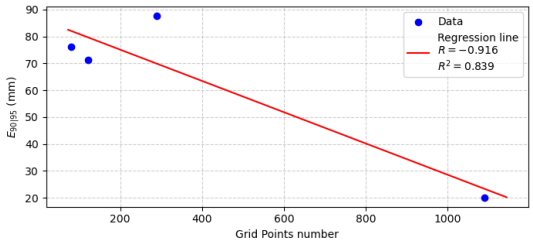
As did in Subsection-6.4.1, the results are presented in terms of $E_{90|95}$:

Radius [mm]	1089 points	289 points	121 points	81 points
$E_{90\|95\,RWP}$	116.84	152.40	174.67	176.21
$E_{90\|95\,RIA}$	20.07	87.64	71.27	76.04
$E_{90\|95\,NRWP}$	80.27	104.95	134.06	116.46
$E_{90\|95\,NRIA}$	45.08	84.04	50.49	64.05
$E_{90\|95\,CWP}$	70.70	117.50	130.81	143.92
$E_{90\|95\,CIA}$	20.83	84.28	64.24	60.97

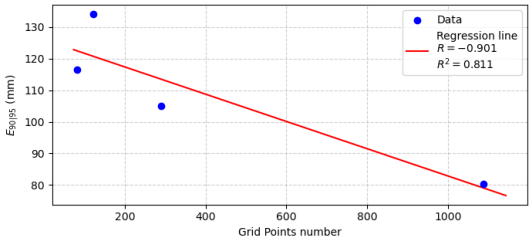
Table 10: Grid point reduction results (2)



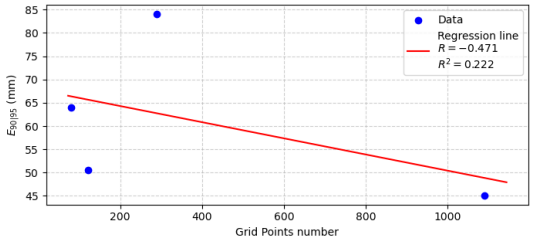
$E_{90|95\,RWP}$ regression



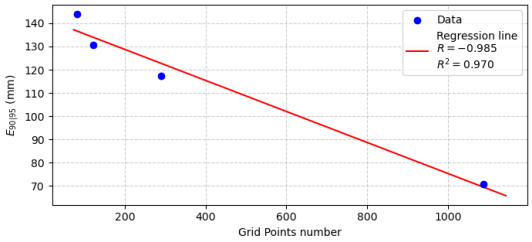
$E_{90|95\,RIA}$ regression



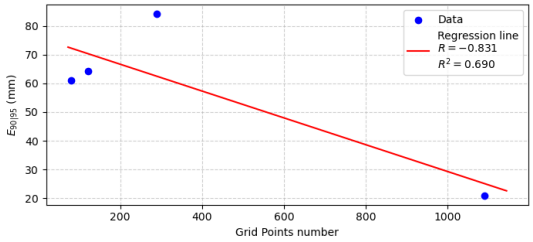
$E_{90|95\,NRWP}$ regression



$E_{90|95\,NRIA}$ regression



$E_{90|95\,CWP}$ regression



$E_{90|95\,CIA}$ regression

Again the increase in $E_{90|95}$ is almost linear with grid points reduction, and the slope is less than the bisector of second quadrant, but with this type of resampling the radius increase is less than previous case, so we can conclude that removing grid rows and columns is preferable in terms of localization accuracy and precision.

6.5 Hsu-Nielsen Impact Source

Device (named after developer of the technique) is an aid to simulate an acoustic emission event using the fracture of a brittle graphite lead in a suitable fitting. This test consists of breaking a 0.5 mm mm diameter and H2 hardness pencil lead approximately 3 mm ($\pm$ 0.5 mm) from its tip by pressing it against the surface of the piece. This generates an intense acoustic signal, quite similar to a natural AE source resulting from the onset of a flaw, that the sensors detect as a strong burst. The purpose of this test is twofold. First, it ensures that the transducers are in good acoustic contact with the part being monitored. Second, it checks the accuracy of the source location setup. This last purpose involves indirectly determining the actual value of the acoustic wavespeed for the object being monitored.

In this section the test data consists in a set of 4 Hsu-Nielsen impact source over the 50 test point seen before, the procedure implemented and the tuning parameters are exactly the same as Section-6.1, 6.2 and 6.3.
The goal now is to test the accuracy and precision in localization changing the AE source compared to the grid node one.

In the following tables the acronyms refer to:

- AWP: accuracy (mean error) in localization over the whole plate
- AIA: accuracy (mean error) in localization in the internal area
- PWP: precision (error standard deviation) over the whole plate
- PIA: precision (error standard deviation) in the internal area
- $E_{90|95,\ WP}$: radius of the area associated with 95% confidence of localization over the whole plate
- $E_{90|95,\ IA}$: radius of the area associated with 95% confidence of localization in the internal area

Localization using standard ΔT method approach (Section-6.1), without sensors pairs region of influence:

Indicator	Value [mm]
AWP	60.80
AIA	34.68
PWP	51.54
PIA	19.01
$E_{90\|95,\ WP}$	106.12
$E_{90\|95,\ IA}$	52.82

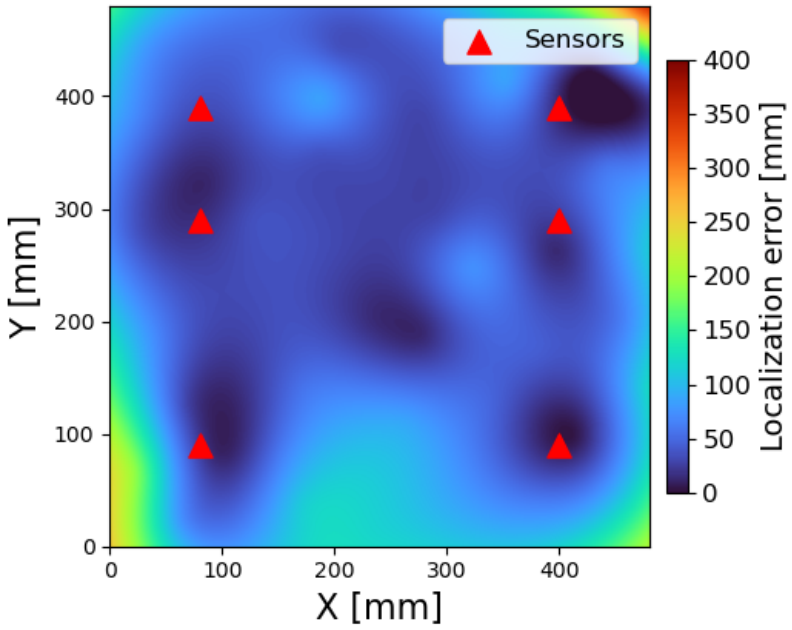


Figure 65: Interpolated localization error heat map for standard ΔT method approach with H-N source

Localization using ΔT method with sensors pairs region of influence (Section-6.2):

Indicator	Value [mm]
AWP	68.12
AIA	34.66
PWP	54.89
PIA	20.52
$E_{90\|95,\ WP}$	128.52
$E_{90\|95,\ IA}$	60.34

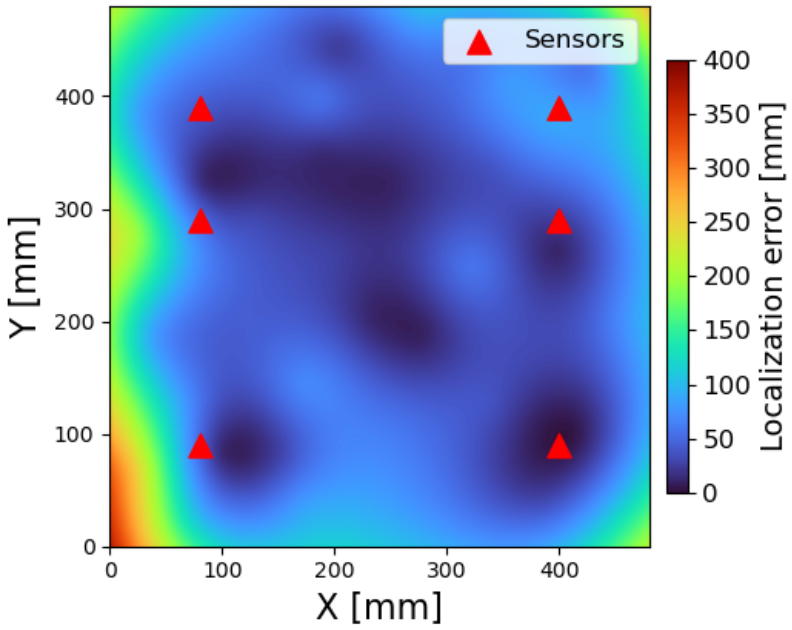


Figure 66: Interpolated localization error heat map for region of influence approach with H-N source

Localization using ΔT method with combined logic (Section-6.3):

Indicator	Value [mm]
AWP	61.33
AIA	34.45
PWP	50.22
PIA	22.82
$E_{90\|95,\ WP}$	110.45
$E_{90\|95,\ IA}$	74.23

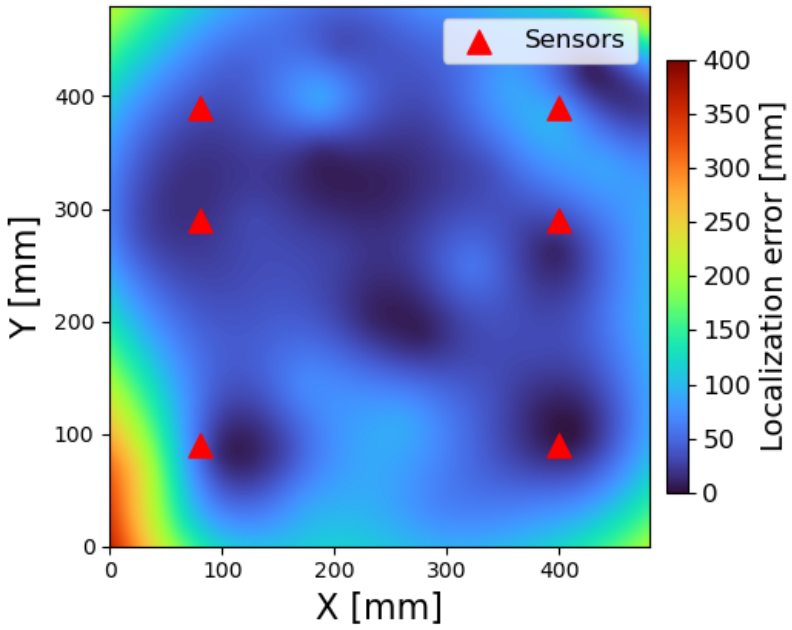


Figure 67: Interpolated localization error heat map for combined approach with H-N source

6.6 Random Forest Application

The main purpose of this section si to implement a machine learning approach to impact localization that reaches at least the performances of the previous ΔT algorithms. In particular, the machine learning algorithm will be the random forest, provided by Python native library `sklearn.ensemble` under the name `RandomForestRegressor`.

The first step is to select and provide to the algorithm the training data, represented by a matrix which each row incorporate the signal features related to a single grid point so it will be a $1089 \times N$ matrix (N number of features).
The signal features selected are:

- Time of Arrival: the time that the signal used to reach a specific sensor, computed using the same methodology reported in Section-3.2.3. Having 6 different sensors and 4 redundancy on impacts, the total number of TOA registered per point will be 24.
- Wave peak: the maximum amplitude reached by the wave signal (24 parameters per grid point),
- Rise time: time elapsed between exceeding of 10% and 90% of the maximum amplitude by the wave signal (24 parameters per grid point),
- Difference in TOA (ΔT): is the parameter on which is based the ΔT method discussed, represent the difference in time of arrival between sensors, so there will be 15 possible difference for 4 different impacts (60 parameters) plus the average value over the 4 redundancy (15 parameters more), reaching a total of 75 different values per grid point.

The signal used in feature extraction is already filtered in the same way reported in the Section-5.

The final training matrix will have shape 1089×147 and as training output the vector of X and Y coordinate per grid node is provided.

The Random Forest algorithm compute the coordinate X and Y simultaneously as dependent parameters, the tuning settings are:

- `n_estimator`=84: number of decision trees embed in the forest, the final result will be the mean of all the single trees' results. There is no risk of overfitting related to the number of tree, increase the amount means that model will generally become more stable, robust and accurate, because you average over a larger number of "opinions", however, beyond a certain threshold the improvements flatten out, while the computation time and required memory increase linearly,
- `max_depth`=11: maximum depth of single tree, defines the maximum number of levels (sequential questions) that any single tree can reach before giving a final answer. By increasing this number the trees will become very deep and complex, the model will learn the training data almost perfectly (overfitting), also capturing the noise or randomness, but will lose the ability to make good predictions on new data,
- `min_samples_split`=5: indicates the minimum number of samples (rows of training/testing dataset) that must be present in a node for the tree to decide to further divide it into two new branches. By increasing it we force the tree to stop sooner to prevent the model from creating overly specific rules based on very little data (helps avoid overfitting),
- `min_samples_leaf`=2: it is the minimum number of samples that must necessarily end up in a "leaf" (the final node that provide the prediction). Increasing this number can have a positive effect on reducing the sensibility to outliers,
- `max_features`= '`sqrt`': when a tree has to decide how to divide a node, it does not look at all the variables available (147 features), but draws a random subset, in this case equal to the square root (~ 12). Increasing the number of variable for example using '`none`' (use all 147 feature available) the trees in the forest will end up all looking alike, because they will always choose the same dominant features to make the first divisions. The forest loses its "diversity" and can becomes less effective.

With this settings we report the results obtained in terms of accuracy, precision and $E_{90|95}$, using the same nomenclature of symbols as Section-6.5:

Indicator	Value [mm]
AWP	34.17
AIA	23.63
PWP	29.63
PIA	18.31
$E_{90\|95,\ WP}$	68.19
$E_{90\|95,\ IA}$	30.80

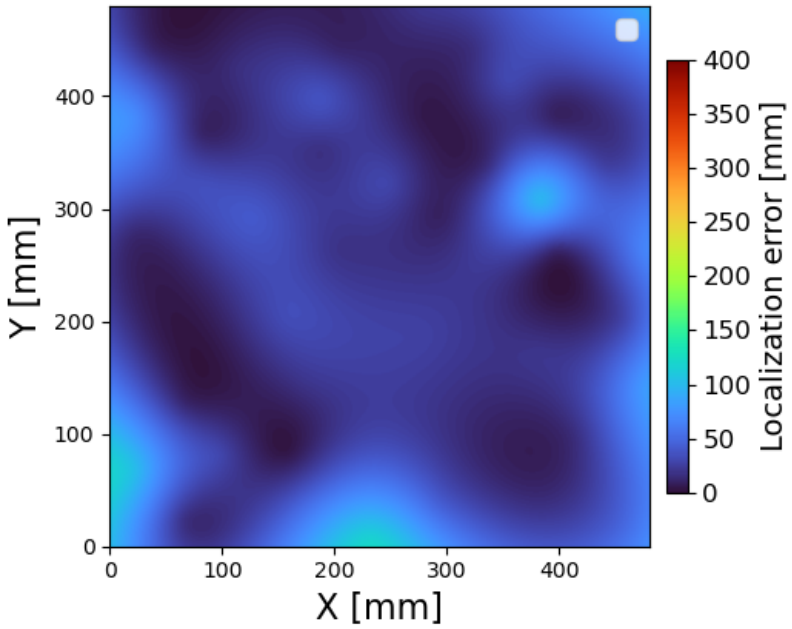


Figure 68: Interpolated localization error heat map for for random forest approach

The results present a slight improvement compared to the combined logic described in Section-6.3, but it is essential to evaluate the change in the precision indicators by reducing the number of training (grid) points, the procedure exploit in Section-6.4.2 is taken as reference:

Indicator [mm]	1089 points	289 points	121 points	81 points
AWP	34.17	49.21	61.49	65.23
AIA	23.63	37.03	37.08	44.30
PWP	32.18	38.56	48.21	44.11
PIA	18.31	33.30	27.69	19.10
$E_{90\|95,\ WP}$	68.19	88.85	125.48	111.66
$E_{90\|95,\ IA}$	30.80	47.98	55.58	67.19

Table 11: Grid point reduction results using Random forest approach

6.7 ΔT Method and Random Forest Comparison

Following the results obtained in this chapter, it is now possible to carry out a comparison between the two methods. The first comparison will be on the localization error of the various test points.

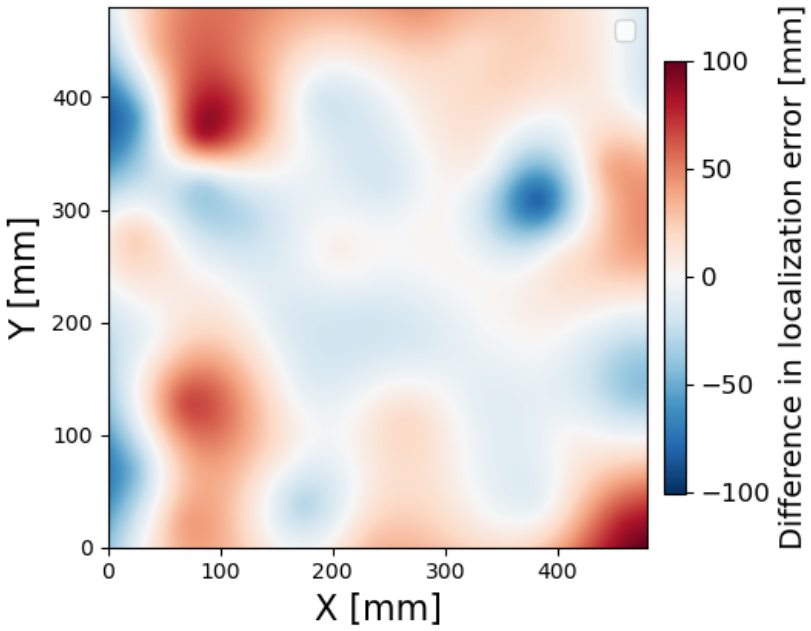


Figure 69: Interpolated heat map of difference in localization error between random forest and combined logic approaches

The map shown in Fig-69 represents the interpolation of the difference in localization error between the two approaches, in particular it is obtained by subtracting the error values associated with the random forest from those of the combined logic. Looking to the map it can be seen that in the areas where the difference is negative the combine logic is preferable, contrarily in the areas with positive values the machine learning approach is advantaged.
As might be expected, the combined approach has a slight advantage in the internal area, thanks to the tuning philosophy applied in Section-6.2.

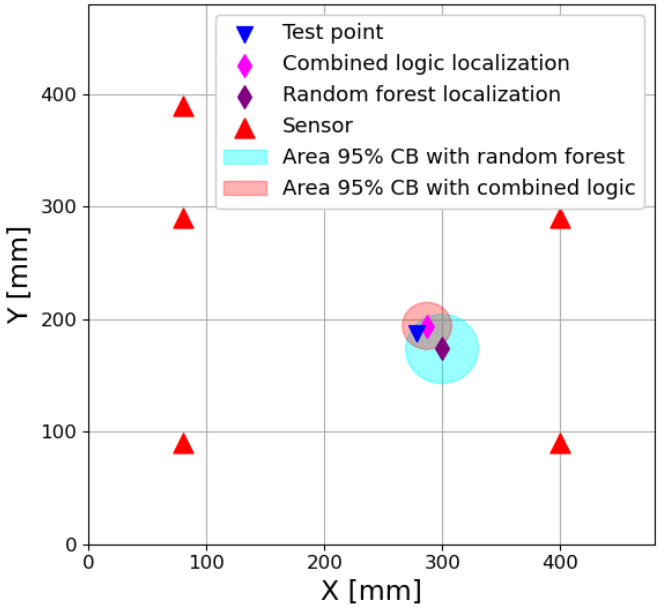
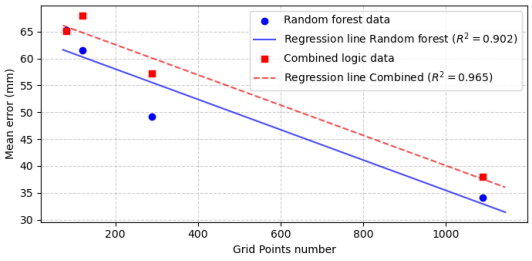
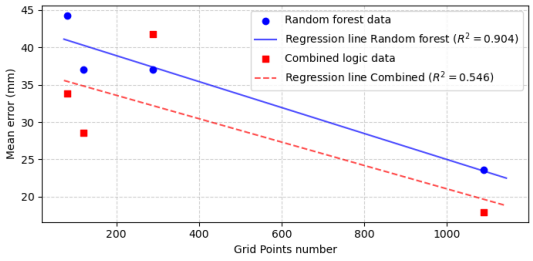


Figure 70: Comparison of localization area associated to 95% confidence between random forest and combined logic

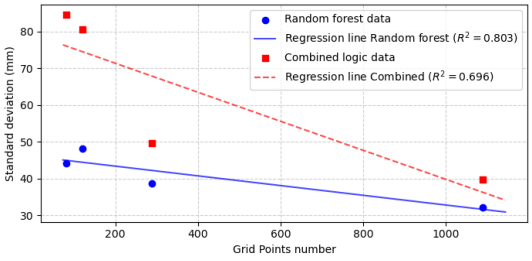
The study continues with the comparison of the performance indices (accuracy, precision and $E_{90|95}$) as the number of grid points varies. The regression curves for comparison are reported.



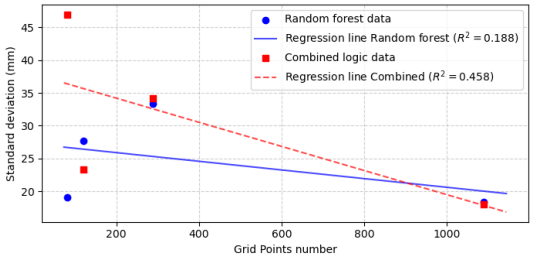
AWP regression comparison



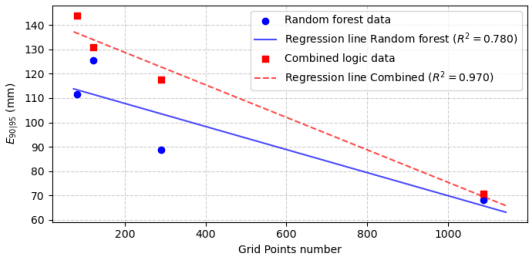
AIA regression comparison



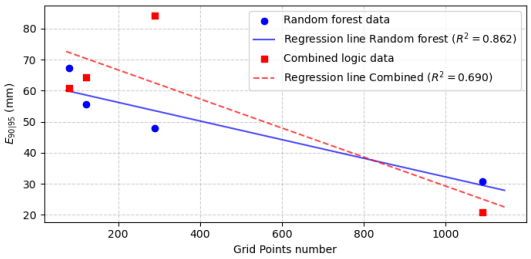
PWP regression comparison



PIA regression comparison



$E_{90|95\,WP}$ regression comparison



$E_{90|95\,IA}$ regression comparison

In this case, by reducing the amount of a priori data, we notice an improvement in the $E_{90|95,\ IA}$ by the random forest compared to the combined logic in the internal zone, although the average error (AIA) remains still lower on average for the latter. This phenomenon is caused by the significant increase in the error's standard deviation (PIA), which contributes significantly in the calculation of $E_{90|95,\ IA}$.

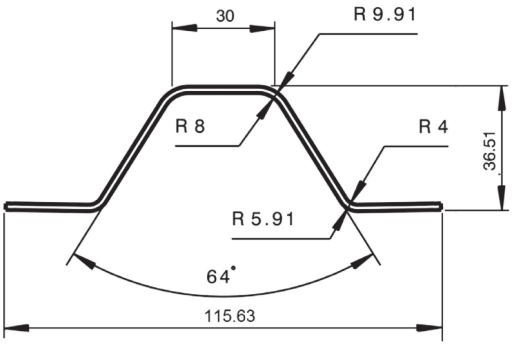
6.8 Results Discussion

As stated in Section-5.2 we are interested in identifying the source with a precision of at least 20 $mm \times 20$ mm. Observing the results we can conclude that the combined approach achieves the objective with a localization radius, associated to 95% confidence, equal to 20.83 mm in the internal area. Considering the entire plate, the localization area is still far from the objective, but the result obtained remains valid as a starting point for further future refinement.
The developed method is unfortunately unstable when the number of grid points is greatly reduced, in this case the random forest approach is preferable.

7 Futures Perspectives

During the experimental campaign conduced in DLR, a second test article was analyzed. It consists of a CFRP plate, nominally the manufacturing process and dimensions are the same as the first plate described in Section-6, the main difference is the presence of an Omega stringer bonded to the bottom of the plate.
The stringer was separately manufactured using M21 / 34% / UD194 / T700 / IMA-12K from Hexply®, it was built also in a quasi-isotropic layup $[-45\ /\ 0\ /\ 90\ /45\ /\ 90\ /\ -45]_S$, the nominal thickness was 1.5 mm with a ply thickness of 0.125 mm. The stringer and the plate were bonded using Loctite Hysol 9466, the adhesive was cured in vacuum at room temperature [38].

(a)



(b)

Figure 74: Plate II bottom view (a), Omega stringer cross-section geometry in mm (b)

The additional complexity presents an intricate challenge in AE source localization, but it is of particular importance because it represents an approach to the structures used in reality on airplanes.
The application and tuning of localization methods discussed during the thesis are

reserved for future applications with the aim of achieving results compatible with the degree of precision required by modern SHM practices.

8 Summary

In the modern scenario of the aerospace industry, the costs of maintenance works and the related losses connected to the downtime of the aircraft constitute a significant expense. To try to reduce these costs, the SHM system is being developed.
The definition of Structural Health Monitoring (SHM) is a system, which detect the various kind of damage and failure induced in the structure, also interpretation and assessment of the damage induced in the engineering structure in order to improve its reliability. We can summarize the objectives of the SHM in two fundamental points: the first is to allow optimal use of the structure while avoiding catastrophic failures, the second is to optimize the organization of maintenance services by reducing downtime.
Within SHM field the Acoustic Emission (AE) technique is frequently used for non-destructive monitoring. It is based on the detection of elastic waves generated by the release of energy within a material, useful for identifying and locating damage or dynamic events in structures. The system consists of a network of sensors glued or co-cured in the structure's area of interest. These sensors capture the waves, convert them into digital format making them available for real time or near real time analysis. There are various techniques for the localization of acoustic emissions such as triangulation or beamforming, but attention is focused on the ΔT method. This method is based on the calculation of the differences in the arrival time of the wave between the sensors. For structures made of isotropic material such as steel or aluminum it provides millimeter level accuracy, while on CFRP structures the performance degrades considerably, reaching a centimeter scale. Since the use of composite structures is increasingly frequent in the aerospace sector thanks to their high stiffness to weight ratio, it is therefore necessary to improve localization methods in order to achieve more precise and reliable monitoring.
The aim of this thesis is to improve the localization accuracy of the already existing ΔT method, then to develop a methodology based on machine learning in order to locate the source within an area of approximately $20 \times 20\ mm^2$.
The experimental campaign was conducted in the Braunschweig headquarters of the German national aerospace research center (DLR) and consists of the analysis of two square plates without defects: one in aluminum with 5 glued piezoelectric sensors and one CFRP with 6 sensors mounted.
On the isotropic aluminum plate, only the ΔT method was tested and a localization accuracy of 5.61 mm was achieved over the whole plate. For the CFRP plate, the results of the ΔT method are divided into results referring to the entire plate and referring to the internal area enclosed by the sensors. Considering the whole plate, the impacts are localized within an area with a radius of 70.70 mm with 95% confidence and an average error of 38 mm, if we consider only the area enclosed by the sensors the radius goes to 20.83 mm and the average error drops to 17.92 mm. The results of random forest machine learning algorithm are only referred to the entire plate and the results obtained are 34.17 mm of accuracy and 68.19 mm of radius.
The set objective is therefore accomplished in the internal area, while externally the results are valid as a basis for future refinement. As a future perspective, we will aim to implement the methods developed on structures with a higher degree of complexity.

9 Appendix

9.1 Iterative weighted centroid

This procedure aim to locate the centroid related to a cluster of point representing the various intersection of iso-ΔT curves. The algorithm is based on the Weiszfeld [37], and is explained as follows:

1. Compute the first centroid location as the mean value of cluster point coordinates
2. Evaluate the euclidean distance between the centroid and each point:

$$d_i = \sqrt{(x_i - x_{centroid})^2 + (y_i - y_{centroid})^2} \tag{55}$$

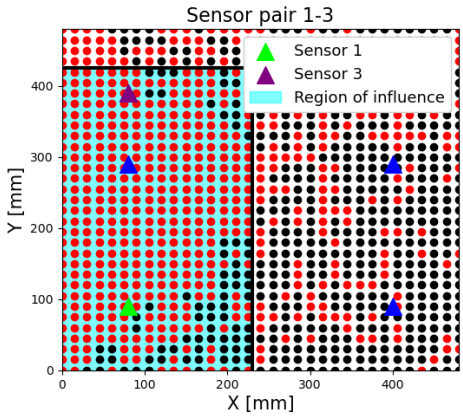
3. Define the inverse of the distance as weight w_i for each point
4. Evaluate the new centroid as:

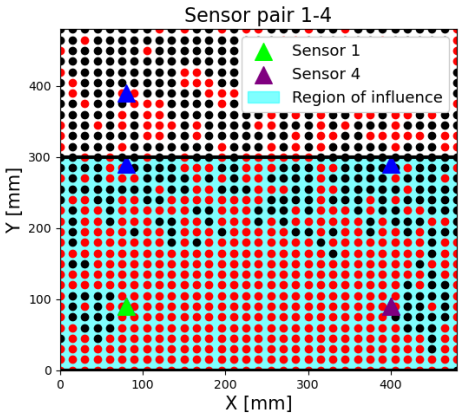
$$x_c = \frac{\sum(w_i \cdot x_i)}{\sum w_i} \qquad y_c = \frac{\sum(w_i \cdot y_i)}{\sum w_i} \tag{56}$$

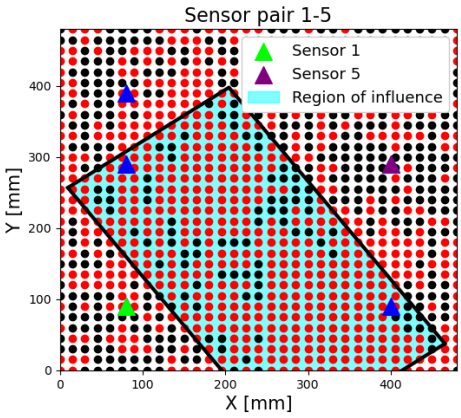
5. Iterate point 1 to 4 until the distance between old and new centroid is less than $\epsilon = 10^{-6}$ m (tuning parameter).

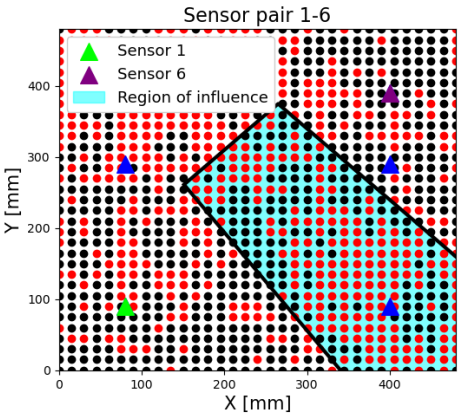
9.2 Sensors pairs region of influence configurations

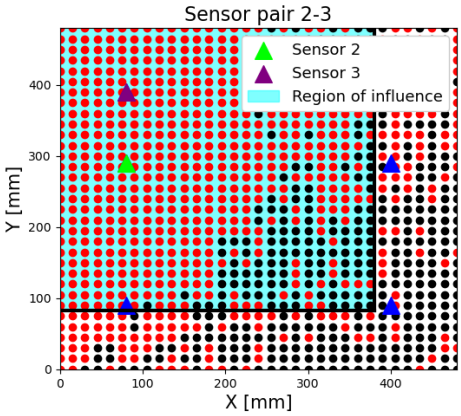
This section contains images showing the configurations of all the sensors pairs region of influence except for sensor pair 1-2 already shown in Section-6.2:

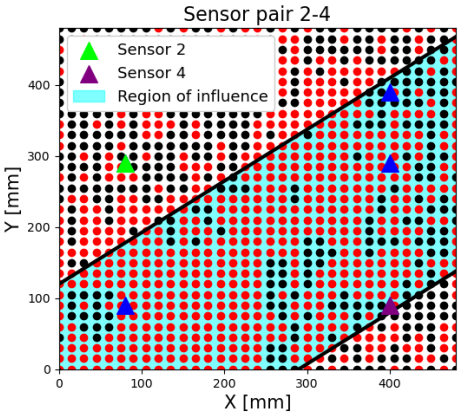


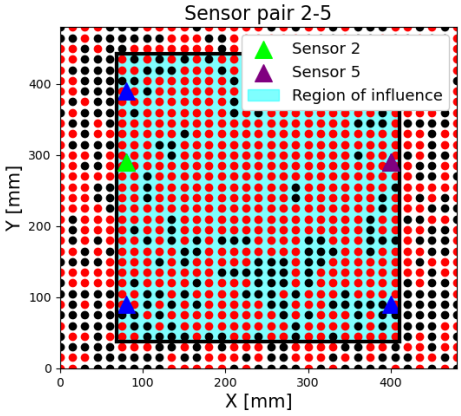


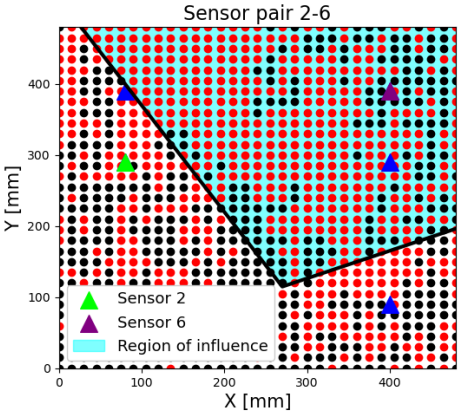

Sensor pair 1-5
Sensor 1
Sensor 5
Region of influence
Y [mm]
X [mm]
0
100
200
300
400

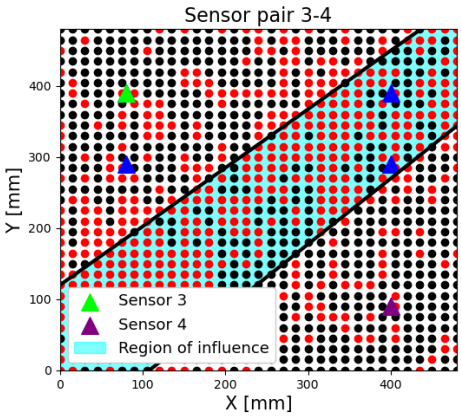

Sensor pair 1-6
Sensor 1
Sensor 6
Region of influence
Y [mm]
X [mm]
0
100
200
300
400

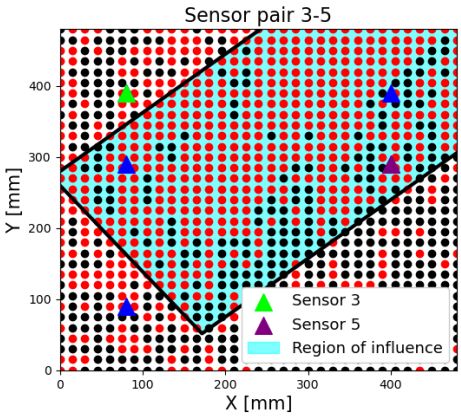

Sensor pair 2-3
Sensor 2
Sensor 3
Region of influence
Y [mm]
X [mm]
0
100
200
300
400

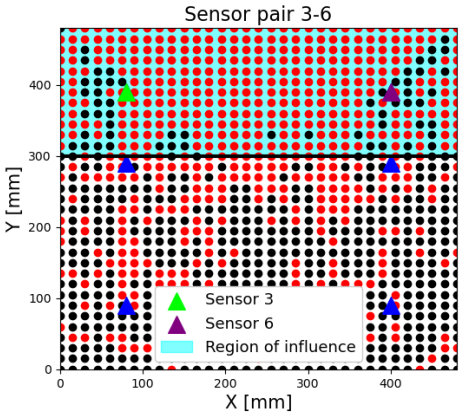

Sensor pair 2-4
Sensor 2
Sensor 4
Region of influence
Y [mm]
X [mm]
0
100
200
300
400

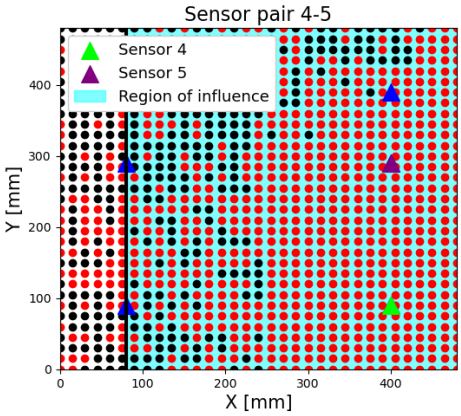

Sensor pair 2-5
Sensor 2
Sensor 5
Region of influence
Y [mm]
X [mm]
0
100
200
300
400

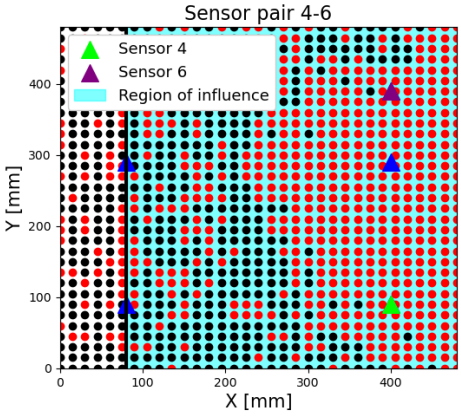

Sensor pair 2-6
Sensor 2
Sensor 6
Region of influence
Y [mm]
X [mm]
0
100
200
300
400

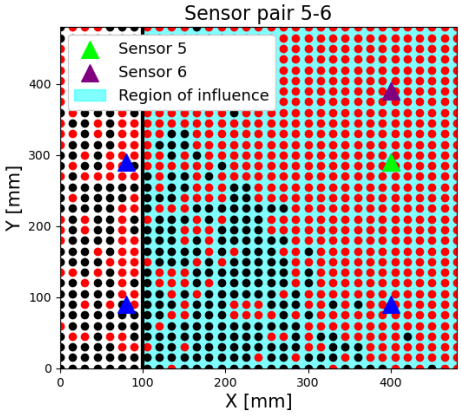

Sensor pair 3-4
Sensor 3
Sensor 4
Region of influence
Y [mm]
X [mm]


Sensor pair 3-5
Sensor 3
Sensor 5
Region of influence
Y [mm]
X [mm]


Sensor pair 3-6
Sensor 3
Sensor 6
Region of influence
Y [mm]
X [mm]


Sensor pair 4-5
Sensor 4
Sensor 5
Region of influence
Y [mm]
X [mm]


Sensor pair 4-6
Sensor 4
Sensor 6
Region of influence
Y [mm]
X [mm]


Sensor pair 5-6
Sensor 5
Sensor 6
Region of influence
Y [mm]
X [mm]

9.3 Overview of results obtained in Subsection-6.4.1 and 6.4.2

This section summarizes the results obtained in Subsection-6.4.1 and Subsection-6.4.2 in terms of Accuracy (mean error) and precision (error standard deviation). In the following tables the symbols shown refer to:

- A_{RWP}: Accuracy in localization considering the region of influence approach over the whole plate
- A_{RIA}: Accuracy in localization considering the region of influence approach over the internal area
- A_{NRWP}: Accuracy in localization considering considering the standard approach over the whole plate
- A_{NRIA}: Accuracy in localization considering considering the standard approach over the internal area
- A_{CWP}: Accuracy in localization considering considering combined logic over the whole plate
- A_{CIA}: Accuracy in localization considering considering the combined logic over the internal area
- P_{RWP}: Precision in localization considering the region of influence approach over the whole plate
- P_{RIA}: Precision in localization considering the region of influence approach over the internal area
- P_{NRWP}: Precision in localization considering considering the standard approach over the whole plate
- P_{NRIA}: Precision in localization considering considering the standard approach over the internal area
- P_{CWP}: Precision in localization considering considering combined logic over the whole plate
- P_{CIA}: Precision in localization considering considering the combined logic over the internal area

Indicator [mm]	1089 points	545 points	363 points	273 points	218 points	182 points
A_{RWP}	64.25	54.6	70.6	56.76	70.15	67.21
A_{RIA}	16.20	23.7	26.04	26.49	20.45	35.41
A_{NRWP}	41.47	49.6	63.27	52.81	51.95	59.61
A_{NRIA}	24.53	28.9	29.85	29.3	31.11	33.7
A_{CWP}	38.0	47.16	67.96	50.88	58.5	61.79
A_{CIA}	17.92	30.02	30.48	30.6	34.9	34.78
P_{RWP}	57.61	54.45	66.76	56.6	77.05	62.9
P_{RIA}	14.51	23.6	31.2	24.6	19.28	36.18
P_{NRWP}	40.50	54.49	72.18	26.3	57.46	62.46
P_{NRIA}	20.83	39.4	36.12	51.4	48.37	47.2
P_{CWP}	39.74	52.7	72.75	53.4	69.12	61.3
P_{CIA}	17.99	42.3	36.39	25.6	49.48	36.46

Table 12: Grid point reduction results, continuation of Table-9

Indicator [mm]	1089 points	289 points	121 points	81 points
A_{RWP}	64.25	70.28	74.91	75.77
A_{RIA}	16.20	40.42	33.95	38.89
A_{NRWP}	41.47	51.6	66.67	55.39
A_{NRIA}	24.53	35.33	26.37	28.64
A_{CWP}	38.0	57.24	68.03	65.07
A_{CIA}	17.92	41.77	28.61	33.82
P_{RWP}	57.61	57.1	68.52	99.54
P_{RIA}	14.51	32.53	32.05	53.84
P_{NRWP}	40.50	47.05	81.16	75.57
P_{NRIA}	20.83	29.04	21.55	34.69
P_{CWP}	39.74	49.64	80.54	84.62
P_{CIA}	17.99	34.17	23.29	46.96

Table 13: Grid point reduction results, continuation of Table-10

References

[1] R. Balaji and M. Sasikumar,"Structural Health Monitoring (SHM) System for Polymer Composites: A Review", *Introduction*, 2016, pp.2.

[2] Daniel Balageas, Claus-Peter Fritzen, Alfredo Güemes,"Structural Health Monitoring", *Introduction to Structural Health Monitoring*, 2006, pp.13-15.

[3] Daniel Balageas, Claus-Peter Fritzen, Alfredo Güemes,"Structural Health Monitoring", *Introduction to Structural Health Monitoring*, 2006, pp.27-28.

[4] S. Gholizadeh, Z. Leman and B.T.H.T. Baharudinb,"A review of the application of acoustic emission technique in engineering ", *A comparison of the AE technique and other NDT*, 2015, pp.1086-1087.

[5] S. Gholizadeh, Z. Leman and B.T.H.T. Baharudinb,"A review of the application of acoustic emission technique in engineering ", *Introduction*, 2015, pp.1075.

[6] Drouillard, T. F.,"Acoustic emission: The first half century ", *No. RFP–4875; CONF-9410182–1*, 1994, pp.1.

[7] S. Gholizadeh, Z. Leman and B.T.H.T. Baharudinb,"A review of the application of acoustic emission technique in engineering ", *Overview on acoustic emission waveform parameters*, 2015, pp.1076-1078.

[8] S. Gholizadeh, Z. Leman and B.T.H.T. Baharudinb,"A review of the application of acoustic emission technique in engineering ", *Overview on acoustic emission waveform parameters*, 2015, pp.1076-1078.

[9] Wojciech Sikorski,"Acoustic Emission: Research and Applications",*Hit detection and Determination in AE burst* pp. 1-5, 2013.

[10] Jierula, Alipujiang, Cong Wu, Abudusaimaiti Kali, and Zhixuan Fu,"A Review of Acoustic Emission Source Localization Techniques in Different Dimensions",*Applied Sciences*, 2024, 14, 8684, https://doi.org/10.3390/app14198684.

[11] KANAKARAJ, Swetha; PITCHAIKANI, Deepthika," Experimental Study of Source Localization in Acoustic Emission using Triangulation Method", In: IOP Conference Series: Materials Science and Engineering. IOP Publishing, 2021. p. 012061

[12] McLaskey, G.C.; Glaser, S.D.; Grosse, C.U.," Beamforming array techniques for acoustic emission monitoring of large concrete structures", 2010, 329, 2384–2394.

[13] Tai, J.; He, T.; Pan, Q.," A fast beamforming method to localize an acoustic emission source under unknown wave speed", 2019, 12, 735.

[14] Matthew Geoffrey Baxter, Rhys Pullin, Karen M. Holford, Sam L. Evans," Delta T source location for acoustic emission", ΔT *location methodology* 2006, pp.1515-1516.

[15] Matthew Geoffrey Baxter, Rhys Pullin, Karen M. Holford, Sam L. Evans," Delta T source location for acoustic emission", ΔT *location methodology* 2006, pp.1517-1519.

[16] Al-Jumaili, Safaa Kh,“ Acoustic emission source location in complex structures using full automatic delta T mapping technique”, *Mechanical Systems and Signal Processing* 2016, 72: pp. 513-524.

[17] Zhou, Zhi-Hua.,“Machine learning”, Springer nature, 2021.

[18] Jierula, Alipujiang, Cong Wu, Abudusaimaiti Kali, and Zhixuan Fu,“A Review of Acoustic Emission Source Localization Techniques in Different Dimensions”,*Artificial Neural networks*, 2024, 14, 8684, https://doi.org/10.3390/app14198684.

[19] Kim, K.-R.; Lee, Y.-S.,“Acoustic emission source localization in plate-like structures using least-squares support vector machines with delta t feature”, Mech. Sci. Technol. 2014, 28, 3013–3020.

[20] Fu, T.; Zhang, Z.; Liu, Y.; Leng, J.,“Development of an artificial neural network for source localization using a fiber optic acoustic emission sensor array”, Struct. Health Monit. 2015, 14, 168–177.

[21] Hasan Ahmed Salman, Ali Kalakech, Amani Steiti, "Random Forest Algorithm Overview", Babylonian Journal of Machine Learning, Vol. 2024, pp. 69-79.

[22] Gerard Biau, Erwan Scornet, "A Random Forest Guided Tour", arXiv:1511.05741v, pp. 4-5, 18 November 2015.

[23] Smith, Steven W. "The scientist and engineer's guide to digital signal processing." 1 Jan. 1997,

[24] Adham R Ismail, Shihab S Asfour, "Discrete wavelet transform: a tool in smoothing kinematic data", Journal of Biomechanics, Volume 32, Issue 3, 1999, pp. 317-321, ISSN 0021-9290,

[25] Kschischang, Frank R. "The hilbert transform." University of Toronto 83 (2006): 277.

[26] Leaman, F., Vicuña, C.M. and Clausen, E. Potential of Empirical Mode Decomposition for Hilbert Demodulation of Acoustic Emission Signals in Gearbox Diagnostics. J. Vib. Eng. Technol. 10, 621–637 (2022).

[27] Yu Ruqing, Wang Keming, Cao Zhong, Roth Michael, “Acoustic Emission”,*Non-destructive evaluation (NDE) of composites: acoustic emission (AE).*

[28] R. Hill, E.U. Okoroafor, A.-M. Priston , “Acoustic emission wave propagation model and issues of damage induction in composites”,*Sources of acoustic emission from fibres and composites*,1996.

[29] Kim Chun-Gon, Sharif-Khodaei Zahra, Xiao Wensheng, Liang Dakai, “Impact Localization”, *Advances in dynamic load identification based on data-driven techniques*, 2021

[30] M.G. Baxter, R. Pullin, K.M. Holford, S.L. Evans, T. Delta, “Source location for acoustic emission”,*Mech. Syst. Signal Process*, 21 (2007), pp. 1512-1520

[31] N. Maeda, “A method for reading and checking phase times in auto-processing system of seismic wave data”,*Zisin=Jishin*, 38 (1985), pp. 365-380

[32] C. Barile, C. Casavola, G. Pappalettera, V. Paramsamy Kannan, “Investigation of an improved time of arrival detection method for acoustic Emission signals and its applications to damage characterisation in composite materials”,*Mech. Syst. Signal Pr*, 223 (2025), Article 111906

[33] Sleeman Reinoud, Van Eck Torild, “Robust automatic p-phase picking: an on-line implementation in the analysis of broadband seismogram recordings”,*Phys. Earth Planet. Interiors*, 113 (1) (1999), pp. 265-275

[34] Ricker, Ormsby, Klauder, Butterworth, “ A Choice of Wavelets”, (2014-12-27)

[35] Stephane Mallat, “ A Wavelet Tour of Signal Processing” *The Sparse Way*, (2008)

[36] Dong XIAO, Zahra SHARIF-KHODAEI, M. H. ALIABADI, “Time of Arrival Extraction for Impact Localization on Composite Structures ”, Imperial College London, 2024

[37] H. USTER and R. F. LOVE, “The Convergence of the Weiszfeld Algorithm ”, pp. 443-450 (01/2000)

[38] J. Moll, J. Kathol, C.-P. Fritzen, M. Moix-Bonet, M. Rennoch, M. Koerdt, A.S. Herrmann, M.G.R. Sause, M. Bach, "Open guided waves",*online platform for ultrasonic guided wave measurements, Struct. Health Monit.*, 18 (5–6) (2019) 1903–1914, doi: 10.1177/1475921718817169

[39] C. Annis, E. Bray, H. Hardy, and P. Hoppe, "Nondestructive evaluation system reliability assessment", U.S. Dep. Def. Wright-Patterson AFB Handb. MIL-HDBK-1823A, 2009.