Proceedings on Engineering Sciences
www.pesjournal.net

CLUSTERING OF LOCALIZED ACOUSTIC EMISSION SOURCES BY THE DBSCAN ALGORITHM IN SEPARATORS

Pavel Makarikhin1
Anastasiya Grigorieva
Margarita Maksimenko

Received 21.01.2023.
Accepted 11.06.2023.
UDC – 620.179.17

Keywords:
Acoustic Emission, Location, Defects, Approximate Calculation, Clusterization, Pressure Vessel

A B S T R A C T

In this paper, the method for locating sources of acoustic emission by approximate calculation of potential coordinates using a grid superimposed on the area around the sensors that detected the wave was proposed. Various noises such as electromagnetic pulses and other external noises were removed from input data. The DBSCAN clustering algorithm was applied. The results were compared with the real state of the object under study after additional control. Analytical and practical research points to the possibility of using the presented method as a tool for determining the coordinates of defect development points and wave velocity.

© 2023 Published by Faculty of Engineering

1. INTRODUCTION

When operating objects, it is very important to know their condition in order to avoid emergencies. Objects such as bridge columns, tanks for the transport of liquids, various structures in industry and many others are under constant loads during their use (Balagurin et al., 2020; Rastegaev et al., 2018; Nosov et al., 2017). There is a wide range of non-destructive testing methods that allow monitoring objects without destroying or dismantling it, such as x-ray inspection (Li et al., 2022), ultrasonic method (Smoqi et al., 2023) One of the most popular approaches is the acoustic emission (AE) method due to its accuracy and environmental friendliness. The AE method finds its application in various areas of technological process monitoring. For example, AE method is allowing to predict the structure of the metal during casting (Yaroslavkina et al., 2020), in (Ser’eznov et al., 2020) AE method was applied in aircraft construction, (Makhnutov et al., 2020) applied AE methods to monitor composite fiber failures. The essence of the AE method is to detect the acoustic waves that arise in the object when defects occur during the operation of the object. To perform this method, AE sensors are installed on the investigated area of the object, then during the continuous loading of the object, various defects begin to appear in it: cracks, splits, material crumbling, and others. When they occur and during their growth, defects emit acoustic waves propagating through the object, which are recorded by sensors. Various parameters of the acoustic wave are recorded: amplitude, energy, time of arrival at the sensor, and many others. This is similar to complex systems with non-periodical dynamics (Pyko et al., 2018). Assessing the condition of an object can be very time consuming due to the size of the object under examination, so it is extremely important to reduce the areas that need to be checked by a specialist. Acoustic
emission data sets are usually quite large, but the correlation of some of the parameters allows them to be reduced for more convenient and faster processing. Analysis of the data allows obtaining information about the defects in the object, their size, development progress and location. At the moment, many methods have been developed for analyzing acoustic emission data for various materials. However, there are still no universal automatic methods of analysis that give an accurate result without the participation of a specialist, therefore, this area is relevant to this day.

At the moment, there are quite a lot of different software for analyzing acoustic emission data, however, the operation of their methods is often hidden from users by trade secrets, and the algorithms used are often not updated along with the development of technologies. (Vallen Systeme, Interunis, Diaton, etc.)

Analysis of large areas of the object under study takes a lot of time for specialists, so location and clustering methods are needed to indicate the areas in which defects are most likely to be located.

There are various approaches to source location: frequency and wave mode analysis (Jiao et al., 2008), laser-based reversal time concept (Park et al., 2012), analytical approach (Grigorieva et al., 2022). The result of the work of location methods are areas in which there will be a defect with the highest degree of probability. The location accuracy is influenced by many factors: the structure and shape of the object under study, its condition, the presence of cavities in it, design features, and much more. Different location methods take these factors into account to varying degrees.

After using location methods, various machine learning methods are often applied to the results obtained to determine the type of signal source and predict the further appearance of defects. In recent years, the field of machine learning has been developing rapidly, various approaches, such as neural networks, clustering and classification algorithms, and others, allow you to get more and more accurate results. Clustering and classification algorithms are used to group data from located sources and form many types of defects. To date, there are many approaches to data clustering that differ in data type and clustering algorithm. Different algorithms have different advantages and disadvantages, as well as different requirements for input parameters. In (Feifei et al., 2011) K-means algorithm was used with cluster centers initializing with random uniform distribution to separate different types of AE signals sources. In (Calubrese et al., 2010) two different unsupervised clustering approaches were used to reduce set of data parameters and identify clusters of different AE signals: Kohonen map and principal component analysis. In (Feifei et al., 2012) DBSCAN (Density-based spatial clustering of applications with noise) (Ester et al., 1996) and K-means were applied to specific material 2.25Cr-1Mo experiment data and were used to separate different types of signals and to extract burst cracking signals.

Solving the problem of localization and clustering of AE signals when examining thin-walled cylindrical vessels used for transporting and storing explosive combustible substances is relevant, as it helps to prevent their early failure, downtime, and even a technogenic or environmental disaster.

2. PRIMARY DATA PROCESSING

When fixing the signals of acoustic emission to sensors on thin-walled vessels, the task goes from spatial to flat - to development of a cylindrical object on a plane. For analysis, the size of the object, the coordinates of the sensors placed on it, various signal parameters, the signal arrived on the sensor and others detected by sensors are used. Various design features of object such as hatches, manholes, supports, unions and others are also taken into account. In Vallen, a single propagation velocity is set for all signals of the experiment, which does not take into account the design and material features, therefore, it is not used in the location and clustering algorithms of this work.

Among the unfavorable factors that have the most negative impact on the result of applying the method are noise-like signals that accompany all modes of operation of most industrial equipment, especially dynamically loaded equipment, and therefore noise is an integral part of any AE diagnostic signals. A high level of noise can lead to a failure in the correct operation of AE signal detectors, which is accompanied by: skips in signal registration; errors in calculating the time of their arrival; the appearance of false or displacement of real location events; incorrect assessment of the hazard class of acoustic sources and, in general, an incorrect assessment of the technical condition of dangerous production facilities.(автореферат , цитирование Игоря Анатольевича)

To solve the problem of detecting signals at the noise level and the possibility of recognizing from several, simultaneously acting acoustic sources, a method was chosen that is used by almost all major scientific groups involved in AE methods, namely: filtering (noise suppression) of the recorded signals in order to bring them to an impulse form for estimation by the amplitude threshold method.

In this work, filtering is carried out: by the amplitude threshold, by the number of sensors that recorded the signal, by the electromagnetic attribute. There are also technical limitations of AE sensors that do not capture the low frequency range below 40 kHz. These technical characteristics lead to forced filtering on a frequency basis. This work uses the VS150-RSC. The VS150-RSC is a piezoelectric AE sensor with integrated Vallen
Acoustic emission data obtained during real experiments often contain a large number of various noises (human factor, electro-magnetic oscillations, cosmic rays (Bonvech et al., 2022), etc.), which are also detected by sensors and significantly increase size of data set, thus it is essential to filter useful information from interference before further work (Davydova et al., 2015; Barat et al., 2010). There are various ways to separate signals from noise: Agletdinov et al., 2020; Rastegaev et al., 2020; Sedlak et al., 2009; Van Der Baan et al., 2015. In this paper, the recorded signals were filtered from obvious noises, the wave amplitude of which does not fall within the range of wave values from AE sources from 40 [dB] to 100 [dB], similarly to Balagurin et al (2020). Next, the signals that were recorded by no more than 4 sensors were removed. Considering the error of the equipment, it is impossible to establish any reliable location of AE sources using less than 4 sensors. The choice of this number of fixed sensors is due to the fact that with a smaller number of sensors, it will be impossible to solve the system of equations for finding the location of the signal source (the method with mathematical equations is used later to verify the reliability of the results obtained by the grid overlay method, see paragraph 3). On the other hand, if information from more than 6 sensors is taken into account, then the set of obtained solutions is sometimes truncated to an empty set due to the influence of data from the last distant sensors, and the more distant the sensor, the greater the error in the information it recorded. Thus, during preprocessing, it is necessary to limit ourselves to 4–6 nearest priority sensors. Then, hypothetical electromagnetic pulses were removed from the remaining data, which can also be captured by sensors from external objects and devices during the experiment. These waves do not carry information about the state of the object and therefore also in this case refer to noise. Such data are distinguished by the fact that their fixation by sensors occurs almost simultaneously, since electromagnetic waves propagate at the speed of light. In the method, it was assumed that signals are electromagnetic, in which the difference in arrival at the sensors is less than 0.2 microseconds [μs]. After filtering the data, the method of approximate calculation on the grid was applied.

### 3. GRID OVERLAY LOCATION METHOD

#### 3.1 Problem Statement

When an acoustic wave propagates through an isotropic material, its velocity in all directions is almost the same. To locate the source, it is assumed that the wave is recorded by, at least, 4 sensors and its speed changes insignificantly. It is also assumed that the wave velocity lies within the range of possible values in the investigated object. The range of velocity values arises in view of the division of waves into Lamb and Rayleigh waves, which have different magnitudes. Depending on the distance from the source to the sensor, different waves of the same signal can be recorded (Grigorieva et al., 2022). Since the calculations take place on the development of the object, it is also necessary to consider the possibility of wave propagation through the edges of the plane along the shape of the object.

#### 3.2 Location method

The method is based on the approximate calculation of the possible coordinates of the wave source. Let several sensors record the wave. For first and second sensors in wave arrival order, the maximum distance at which the wave source can potentially be relative to them is calculated.

\[
d_{\text{max}} = \frac{\sqrt{(x_2-x_1)^2+(y_2-y_1)^2}}{2}
\]  

where \(x_0,y_0\) - coordinates of the i-th sensor in order that recorded the wave.

Further, a grid with a certain step is superimposed on the area in which the sensors that recorded the wave are located within this distance. For every point grid, the condition of changing the velocity within the specified accuracy is checked:

\[
|v_1-v_2| < \text{accuracy}
\]

where accuracy is predefined,

\[
\begin{align*}
  v_1 &= \frac{\sqrt{(x_2-x)^2+(y_2-y)^2}+\sqrt{(x_1-x)^2+(y_1-y)^2}}{t_2-t_1} \\
  v_2 &= \frac{\sqrt{(x_3-x)^2+(y_3-y)^2}+\sqrt{(x_2-x)^2+(y_2-y)^2}}{t_3-t_2}
\end{align*}
\]

After calculations for all grid vertices, a set of potential source points for different velocities is formed. Then, for each point of the resulting set, it is checked that among the three closest sensors to the given point there are at least 2 detected corresponding signal, allowing some error in the operation of the sensors.

| Table 1. The set of possible AE source points for a single signal. |
|----------------|----------------|--------|
| x              | y              | speed  |
| 1              | 226.82         | 314.32 | 141.23 |
| 2              | 229.62         | 327.22 | 164.11 |
| 3              | 229.82         | 328.12 | 165.48 |
| 4              | 230.72         | 332.12 | 171.26 |

The Table shows an example of data processing for a signal recorded by 5 sensors and shows the result of the proposed location method. The first and second columns indicate the most probable coordinates of the AE signal...
source, and the third column indicates the probable propagation velocity of this signal.

Figure 1 shows the result of the method for a single acoustic wave: the sensors that detected the wave are highlighted in green, red - set of potential points. At the last stage sets with more than 4 solutions (possible location points) were removed and the average value was calculated for each set of points.

Figure 1. Grid overlay location method result

After averaging the values in each set of points, a data set was obtained for further clustering.

4. DATA CLUSTERIZATION

For data clustering, the DBSCAN algorithm was chosen because it does not require knowing in advance the number of clusters, it gets the same result for the same data set, and it is resistant to outliers.

The clustering algorithm was applied to the previously obtained data set to determine the points clustering areas. The data obtained using the algorithm (Grigorieva et al., 2022) applied to the experimental data were also clustered to check the correctness of the location method.

The essence of the developed clustering method is to automatically calculate the ranges of characteristics characteristic of this particular object for one of the dangerous zones in which there is definitely a defect and, based on this, find the remaining zones of probable defects. For the method to work, the coordinates of the rectangle in which the required zone is located and the number of high-amplitude pulses in this rectangle are required. The number of pulses is determined empirically and varies from 3 to 10 depending on how damaged the object is, based on (Gomera et al., 2014).

Since it is important to take into account not only the location of the points, but also the values of the amplitude and speed, for the correct formation of clusters. The selection of parameters was based on the following articles: Gomera et al., 2014; Yaroslavkina, 2018; Rastegaev et al., 2022.

The following metric was developed and applied:

$$f(i,j) = w_1 \times l + w_2 \times v + w_3 \times a$$ (4)

$i, j$ – points in data set; $w_1, w_2, w_3$ – weight coefficients of distance, speed and amplitude, respectively;

$$l = \sqrt{\left(\sum_{k=0}^{n} x_{ik} - \sum_{k=0}^{m} x_{jk}\right)^2 + \left(\sum_{k=0}^{n} v_{ik} - \sum_{k=0}^{m} v_{jk}\right)^2}$$ (5)

$$v = \left|\sum_{k=0}^{n} v_{ik} - \sum_{k=0}^{m} v_{jk}\right|$$ (6)

$$a = \min\left(\sum_{k=0}^{n} a_{ik}, \sum_{k=0}^{m} a_{jk}\right)$$ (7)

where $x, y, v, a$ – averaged coordinates, speed, amplitude of points in sets which were described in section 3.2., $n, m$ - number of points in this sets.

For each of the parameters (distance, speed and amplitude) the corresponding threshold values were used to calculate the global epsilon used in the DBSCAN algorithm:

$$\text{EPS} = w_1 \times \text{eps}_1 + w_2 \times \text{eps}_v + w_3 \times \text{eps}_a$$ (8)

Weights $w_1, w_2, w_3$ and epsilons $\text{eps}_1, \text{eps}_v, \text{eps}_a$ are automatically calculated for each object of study based on data values. Automatic calculation of parameters for each object is increasing the accuracy of clustering.

After applying the algorithm, clusters with a large number of elements were removed from the resulting set of clusters, since in real cases such clusters are most likely signals received as a result of friction of object against a support that are not defects. Depending on how the experiment was conducted, clusters with 10-20 elements were considered as large.

As a result, the resulting clusters display areas on the object that are most likely to have defects and that should be checked by a specialist.

5. EXPERIMENT

Data from several real experiments were used to test the methods. The formation and development of defects in the process of object loading were controlled by recording acoustic emission signals in real time with an Amsy-5 Vallen system (Germany) as in (Damaskinskaya et al., 2021). The technical features and characteristics of piezoelectronic sensors were given in paragraph 2. Software was developed for processing AE data using the proposed method. It is implemented on a modular basis, i.e., is a sequentially executed scripts of procedures: reading data, converting it, processing it according to the algorithm described above, unloading, clustering and visualization.
scripts are implemented in R, C# and Python programming languages.

5.1 Experiment 1

In the first experiment, the object of the study was a cylindrical pressure vessel: a separator made of steel 09G2S-17+08Kh13 with a height of 3200 [mm], an inner diameter of 2000 [mm] and a wall thickness of 12 [mm]. A hydraulic test was carried out: the vessel was filled with water and been under increasing pressure for 50 minutes. During the loading process, several cracks appeared in the object, which were recorded by specialists. The initial data contained 24579 lines of sensor data, of which 4856 remained after the primary data processing, making up 1061 signals.

5.2 Experiment 2

In the second experiment, the object of the study was a low-pressure separator made of steel 09G2S-12+08Kh13 with a height of 6674 [mm], inner diameter of 2400 [mm], and wall thickness of 14 [mm]. The capacity of the device - 31.0 [m³]. The same hydraulic test as in experiment 1 was carried out. The initial data contained 24579 lines of sensor data, of which 4856 remained after the primary data processing, making up 1061 signals.

5.3 Experiment 3

In the third experiment, the sensors were calibrated, during which AE events were simulated using a special device in different places of the object. By applying this device to the object under study, acoustic waves are simulated, similar to the waves that occur during the occurrence of defects, in order to fine-tune the AE. After applying the location method, 5 groups of AE sources were correctly identified, corresponding to the simulated signals. The initial data contained 431 lines of sensor data, of which 148 remained after the primary data processing, making up 37 signals.

6. RESULT

After applying the location method to first experiment data, the sets of possible AE source coordinates were calculated for 206 signals from the data set. For each of resulting signals, sets as in Figure 1 were received and displayed on a single location diagram in Figure 2.

Signals with 1 solution are shown in green, signals with 2 or 3 solutions – in blue, signals with 4 solutions are shown in red. It is important to note that the speed is calculated for each signal and for each solution individually. Thus, Figure 2 shows all the points of location of AE sources obtained by the method for the object under study. Moreover, the speeds of different AE signals are various. They are allowed to be different even for each of the four solutions of one signal.

Figure 2. Sets of the possible positions of AE sources.

Dot clustering areas coincide with the areas indicated by specialists for additional control, due to the possible presence of defects in these areas. It shows the correctness of the application of this method when examining objects of this type.

In the second experiment, after applying the location method 534 points dataset was obtained. As a result of clustering, 5 clusters were obtained.

In Figure 3, the red rectangles mark the areas in which defects were actually found. Points with an amplitude of 40[dB] to 55[dB] are colored green, those with an amplitude of 55[dB] to 60[dB] are yellow, and red with an amplitude greater than 60[dB]. Points that did not fall into any cluster are colored black. Figure 3 shows that clusters were formed next to each problem area, and several clusters were formed near the area of the structural supports, where many AE signals often occur.

Figure 3. Experiment 1 clusterization result

In Figure 4 clusters of simulated AE sources shown as colored circles. The location of all sources was correctly determined and clustered.
7. CONCLUSION

The results obtained demonstrate that the proposed methods adequately reflect areas containing defects and their development. The method has been tested on the real experiments. The application of the developed method significantly reduces the areas of the object that should be checked by a specialist. In the future, it is planned to optimize the metric function in the DBSCAN algorithm for different types of objects and apply more complex criteria for epsilon.

When developing various methods for working with AE data, it was decided to combine all the results described in this article and the results obtained in previous works into a single code library. The library will allow to apply methods and combine the results more conveniently and faster. The library will contain methods for primary data processing for cleaning them from noise, locating sources of AE signals, and methods for clustering and classifying sources.

To develop the library, the C# programming language was chosen in view of the convenience of working with data, as well as the availability of various packages with functionality for applying machine learning and mathematics methods. At the moment, the problem of combining all the developed methods is being solved in view of the fact that they were written in different programming languages, and therefore it is required to refactor the code written over several years.

Acknowledgement: The authors would like to express gratitude to Gomera Victor for sharing the idea of grid location method and cooperation on acoustic emission issues.

Supported by Saint Petersburg State University, project ID: 93024916

References:


<table>
<thead>
<tr>
<th>Name</th>
<th>Institution</th>
<th>Email</th>
<th>ORCID</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pavel Makarikhin</td>
<td>Saint Petersburg State University, Saint Petersburg, Russia</td>
<td><a href="mailto:St054466@student.spbu.ru">St054466@student.spbu.ru</a></td>
<td>0000-0001-6152-1430</td>
</tr>
<tr>
<td>Anastasiya Grigorieva</td>
<td>Saint Petersburg State University, Saint Petersburg, Russia</td>
<td><a href="mailto:St053387@student.spbu.ru">St053387@student.spbu.ru</a></td>
<td>0000-0002-9267-340X</td>
</tr>
<tr>
<td>Margarita Maksimenko</td>
<td>Saint Petersburg Mining University, Saint Petersburg, Russia</td>
<td><a href="mailto:Maksimenko_mv@pers.spmi.ru">Maksimenko_mv@pers.spmi.ru</a></td>
<td>0000-0002-2588-0310</td>
</tr>
</tbody>
</table>