AUTOMATED METHOD FOR STATISTICAL PROCESSING OF
AE TESTING DATA

V. A. Barat and A. L. Alyakritskiy
Research Dept, Interunis Ltd., bld. 24, corp 3-4, Myasnitskaya str., Moscow, 101000, Russia

Keywords: signal processing, AE data classification

Abstract

Acoustic emission (AE) as a nondestructive testing method allows estimating the condition of sophisticated industrial objects, detecting defects at their initiation, and preventing development of such defects. The reliability of detection and the accuracy of determination of AE source location depend on the correct interpretation of the AE testing results. Testing procedures of different industrial equipment are intended for estimating condition of the objects on the basis of AE parameters, while the AE signals proper are not used in analysis as the primary diagnostic data.

We have developed the statistical method of analysis to study the whole complex of measured data, both the AE parameters and signal waveforms. This method is based on the two-level clustering of data, such as AE signals from separate channels, and "groups of signals" formed on the basis of the primary clustering. The algorithm allows the determination of the quantity of AE sources, and estimates the degree of their danger without resorting to the preliminary location. The zonal location is carried out after the determination of AE sources. It avoids the so-called false locations, which considerably complicate the location map. On the basis of classification results, one can correct the AE signal arrival time, and such correction makes it possible to improve the accuracy of the AE source location.

The characteristic property of the developed method is data processing in an automatic mode, with minimal operator intervention. In this case there is no need to employ a high-skilled operator. Since the AE signals themselves are used for analysis, the obtained results prove to be more reliable than those obtained when working with the AE parameters. The algorithm built on the developed method has been successfully applied for processing data obtained as a result of laboratory research of the corrosion development, for studying reinforced concrete structures, and also as an additional instrument for processing data of the industrial AE testing.

Introduction

Each AE event defines a unique process occurred at a certain point of the test object. When AE signals propagate from the point of emission to a sensor, the waveforms are complicated due to conversions to different modes of waves, multiplicity of distribution paths, and due to wave velocity dispersion by frequency. AE signals contain information not only about the AE source event for this signal, but in the highest degree about parameters of acoustic and electric path. In this connection construction of the analytical or even numerical diagnosis model for interpretation of AE signals appears to be an intricate and nontrivial problem, the solution of which cannot be generalized for the test equipment of different types.

While the data interpretation based on accurate diagnostic models appears to be difficult, the statistical analysis is a reasonable choice for carrying out the data interpretation and classification with the high certainty. At present the statistical methods of data analysis are widely applied both in industrial AE systems and in laboratory research, and the applications are varied: data clustering, correlation analysis, and check of different statistical hypotheses.

In this paper the statistical method of analysis enabling the automatic clustering of AE testing data is presented. This method makes it possible to process bulk data obtained as a result of the AE testing or during the laboratory research in the absence of a priori information. As a consequence of processing, the data are structured and organized; each cluster formed as a result of analysis characterizes an AE source at a definite stage of development.
Method Description

The clustering of AE data is used rather frequently. However, different methods of data processing pursue different aims, and have special features of implementation [4,5,7,8]. The key features of a given method is, firstly, the possibility for data processing in an automatic mode with minimum operator’s involvement and a minimum number of settings, and, secondly, the possibility to carry out the analysis of the heterogeneous diagnosis information.

The initial data for algorithm realizing the given method can be both the AE signals and also their parameters computed under data acquisition in on-line mode. Even if the AE count rates are impossible, more than half the AE signals can be substituted for values of AE parameters (arrival time, rise time, energy, etc.) without loss of the processing accuracy.

![Flowchart of automated method of statistical processing of AE data.](image)

Fig. 1. Flowchart of automated method of statistical processing of AE data.

Figure 1 shows the flowchart of the proposed method. The data are processed in two stages. At the first stage the clustering of AE signals takes place. The data recorded by each measuring channel are analyzed individually. The correlation coefficient is used as a measure of similarity of each pair of signals. The “classes of signals” are formed at this first stage.

Next, the classified signals are combined in groups in such a way that one group will include the signals of different measuring channels, which belong to the same AE event. The next stage of algorithm is the formation of “classes of groups”; to the same “class of groups” assigned are the groups wherein the signals recorded by one and the same “class of signals” correspond to the same “classes of signals”. The quantity and parameters of AE sources are estimated by the results of classification of signal groups.

When a part of AE signals is defined only by parameters, and waveforms are absent, the “classes of signals” and “classes of groups” are formed on the basis of incomplete diagnosis information. For classification of AE signals, specified only by their parameters, each “class of groups” is characterized by a set of features. The classification in this case is accomplished on the basis of multidimensional empirical distribution function built for the calculated features.

Clustering of Acoustic Emission Signals

For AE signal clustering, it is necessary to determine a distance measure [3]. In order that an estimation error of AE parameters does not influence on the classification result, the correlation coefficient $r$ was used as a measure of similarity of two signals; see Eq. (1). The preliminary analysis has shown that signals relevant to the same AE source and recorded by the same
measuring channel have a high correlation coefficient, whose value varies from 0.70 to 0.99 depending on the nature of the AE generating process.

\[
r = \frac{\sum_{i=1}^{N} (x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{N} (x_i - \overline{x})^2} \sqrt{\sum_{i=1}^{N} (y_i - \overline{y})^2}}
\]

(1)

Using the correlation coefficient as the distance measure of AE signals, an automatic correction of arrival times of signals, which belong to the same cluster, can be carried out, because the cross-correlation function reaches its peak at the time corresponding to a difference of arrival times of AE signals.

When processing bulk data, the calculation of correlation coefficients for all pairs of signals requires a long time; to speed up the data processing, a wavelet decomposition may be used, namely, a signal decomposition by wavelet packets [2]. The decomposition by wavelet packets is one of variety of the multi-resolution analysis, and it is the signal decomposition on the basis of wavelets (in a general case by the Riss basis) specified for a sequence of subspaces embedded to each other.

\[
w_{2n}(x) = \sqrt{2} \sum_{k} h(k) w_{n}(2t - k) \quad w_{2n+1}(x) = \sqrt{2} \sum_{k} g(k) w_{n}(2t - k),
\]

(2)

where \( g(k) = (-1)^{k} h(1 - k) \) is defined by the type of wavelet function.

Under a discrete wavelet transform the signal is decomposed into a low-frequency component - approximation and a high-frequency component - detailing. At the next level both the approximation and the detailing are subject to further decomposition. So, at \( j \)-level of decomposition the \( 2^j \) coefficients are calculated. In this case the coefficient \( (j, k) \) localizes energy within the frequency range (Eq. 3), where \( \Omega_0 \) is the frequency corresponding to a half of the sample rate frequency.

\[
\Delta \Omega = \left\{ \frac{\Omega_0}{2^j} k, \frac{\Omega_0}{2^j} (k + 1) \right\}.
\]

(3)

As a rule, the AE signal energy is non-uniformly distributed on the spectrum, and often 3 or 5 coefficients of cluster decomposition are sufficient for location of 95 - 99% of energy. Moreover, the dimension of informational components, as a rule, is at least four times smaller than the original signal dimension. The research has shown that the precise estimate of correlation coefficient based on the weighted sum of correlation coefficients of wavelet-packet decomposition components is 85%.

With the clustering of AE signals, the quantity of clusters is to correspond to the quantity of the assumed AE sources. In designing the clustering algorithm, a particular attention was given to the true determination of clusters quantity. For illustrating this aim, we suggested a two-stage clustering diagram, as shown in Fig. 2; that is, the flowchart of clustering algorithm. At the first stage, a hierarchical agglomerative algorithm with complete link clustering is applied. At the second step an iteration analysis is applied, the distances between centers of mass of the clusters formed are calculated, and two clusters having the maximum and above-threshold distance between the clusters centers are merged. Thereafter, the re-clustering is accomplished by the method of k-averages.

When using this flowchart of clustering algorithm, the clustering is flexible and controllable; it uses two settings – the threshold value for signal addition to the cluster (at hierarchical clustering) and the threshold value for the inter-cluster distance. The setting values are selected on the basis of information about the test object and the testing conditions.

**Clustering of Groups of Acoustic Emission Signals**

At the next stage of algorithm, the groups of signals relevant to the same AE event are analyzed. For formation of such the groups analyzed are the arrival times of AE signals, the differ-
Hierarchical clustering with clusters merging as per the “complete link” rule

Is the minimum distance between centers of clusters less than the threshold one?

Yes

Clusters merging

No

Reclustering by the method of k-average

Fig. 2. Flowchart of two-stage clustering algorithm of AE signals.

ence of recording time of the first and last signals in the group should not exceed the specified value, which is determined by the ratio of its maximum overall dimension and minimum velocity of acoustic wave propagation experimentally measured for the given test object. An AE signal beyond the time window is initial AE signal for the next group.

To the same “class of groups” assigned are the groups wherein the signals recorded by the same measuring channels corresponding to the same classes of signals. The study of algorithm has shown that for assigning the group of signals to one or other class it is sufficient that the classes of signals coincide at least for two channels in the group.

The membership of “classes of groups” can be also defined for AE signals, even with no information about the AE signal waveform. For this purpose, each “class of groups” obtained is characterized by the values of features listed below representing a median of distribution for each class.

− Numbers of three channels with minimum time of signal arrival \{n_{t1}, n_{t2}, n_{t3}\}
− Numbers of three channels with maximum signal amplitude \{n_{a1}, n_{a2}, n_{a3}\}
− Maximum value of signal amplitude in the group \(A_{max}\)
− Ratio of amplitudes of signals recorded by channels \(n_{t2}\) and \(n_{t1}\), and also \(n_{t3}\) and \(n_{t1}\) \(\{A_{21} & A_{31}\}\)

The signal group belonging to one or other class is defined similarly to the k-averages method, by the minimum distance between the features, characterizing each “class of groups” and the group of features to be analyzed.

The “classes of groups” obtained by such a manner characterize the potential AE sources with a high degree of probability. The quantity of “classes of groups” corresponds to the quantity of AE sources; the numbers of channels, by which the signals with the minimum arrival time are recorded, define the number of location zone. The quantity of groups in each class and the energy of AE signals included into the group can define the degree of danger of AE source.

Results and Discussion

The practical examples can visually confirm the effectiveness of the present method. To research regularities of AE in concrete, a series of experiments was carried out. On a concrete cube with a 1-m side length, eight AE sensors were placed, one at each vertex. The AE sensors were used both for AE measurement mode and as a pulser for the simulation of AE sources.

To illustrate advantages of the present method, two of the experiments were selected. Experiment 1 illustrates the case when the AE sources are not found because of mechanical noise and impacts. There are two AE sources, a pulser that simulates a growing defect, and a hammer that simulates mechanical noise of equipment. Figure 3a shows the result of volume location of
the AE sources. Indications distributed over the whole field of location are generated by the mechanical noise source or hammer blows. As the hammer impacts are made manually, the relevant signals have different waveforms. In this case, determination of the arrival time by the threshold method entails the difference in arrival time errors. The errors of arrival time determination differ in their meaning and sign for the signals with different waveforms. This is one of the possible reasons for the distribution of source locations.

Application of the statistical method of data analysis allows not only forming “the class of groups” of AE impulses emitted by the simulator, but also pointing out as an independent class the signals relevant to the hammer blows. Table 1a shows the results of operation; the attributes given in this table conform to the list of “class of group” features. The names of features in Table 1 comply with the previous list of features. Each line of the table characterizes one "class of groups".

Table 1a AE class data – 2 classes

<table>
<thead>
<tr>
<th>Class No.</th>
<th>Qty of Class Elements</th>
<th>A\textsubscript{max}</th>
<th>n\textsubscript{t1}</th>
<th>n\textsubscript{t2}</th>
<th>n\textsubscript{t3}</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>47</td>
<td>7805</td>
<td>2</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>2</td>
<td>55</td>
<td>20480</td>
<td>6</td>
<td>7</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 1b AE class data – 6 classes

<table>
<thead>
<tr>
<th>Class No.</th>
<th>Qty of Class Elements</th>
<th>A\textsubscript{max}</th>
<th>n\textsubscript{t1}</th>
<th>n\textsubscript{t2}</th>
<th>n\textsubscript{t3}</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>24</td>
<td>1980</td>
<td>6</td>
<td>7</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>23</td>
<td>2000</td>
<td>6</td>
<td>7</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>24</td>
<td>5593</td>
<td>8</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>4</td>
<td>26</td>
<td>2195</td>
<td>8</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>5</td>
<td>24</td>
<td>3659</td>
<td>1</td>
<td>7</td>
<td>3</td>
</tr>
<tr>
<td>6</td>
<td>22</td>
<td>6393</td>
<td>1</td>
<td>7</td>
<td>3</td>
</tr>
</tbody>
</table>

In Experiment 2, there are six AE sources in all. Two types of acoustic waves were generated at three positions on the concrete cube surface by means of a pulser and a Hsu-Nielsen source. Figure 3b shows the three locations of the AE sources. During the statistical processing of data, application of the present method can give six classes of events corresponding to two different sources of AE at three points on the test object surface, as shown in Table 1b. There are two "classes of groups" in each location zone, which is defined by the channels number \{n\textsubscript{t1}, n\textsubscript{t2}, n\textsubscript{t3}\}.

One more useful implementation of the present method is structuring and data compression. The data structuring is realized due to replacement of the AE signals being analyzed by the "classes of signals" and "classes of groups". Thus, the quantity of the information under analysis is reduced considerably. For example, during experimental AE studies from pitting corrosion growth, several hundred thousand signals were recorded; after the statistical analysis of data in an auto-mode without pre-processing about 20 representative classes of "groups of pulses" describing different stages of corrosion damage development were defined [6].

In practical AE applications, it is effective to use automated method of statistical analysis for processing of the AE monitoring data, especially when it is necessary to analyze changes in the
structure of AE signals recorded for any length of time, and also in the case of testing sophisticated industrial facilities, for which the construction of an acceptable location scheme appears to be difficult.

**Conclusion**

In this paper described is an automated statistical analysis method of AE data, which makes it possible to structure the AE test data through partitioning into different clusters or the groups of signals, characterizing the different AE sources. The key features of the given method are, firstly, the possibility to process data in an automatic mode with minimum involvement of operator and a minimum number of settings, and, secondly, the possibility to carry out the analysis of the heterogeneous diagnosis information. Based on the results of statistical analysis, it is possible to specify the quantity of AE sources, to carry out its zone location, and to get additional evaluation of the danger criterion without resorting to the preliminary location.

When using the correlation coefficient as a measure of proximity under cluster analysis of AE signals, we can carry out an automatic correction of signal arrival times belonging to the same clusters, because the cross-correlation function reaches its peak at the point of time corresponding to the difference of arrival times of AE signals.

**References**