Evaluation of Increasing Damage Severity in Concrete Structures by Cluster Analysis of Acoustic Emission Signals

Giuseppe CAMPANELLA, Luigi CALABRESE, Edoardo PROVERBIO
University of Messina, Dept. Industrial Chemistry and Material Engineering, Viale F. Stagno d’Alcontres 31, S. Agata di Messina, 98166 Messina, Italy; placido.campANELLA@tin.it, calabrese@ingegneria.unime.it, proverbi@ingegneria.unime.it, tel.+39 3977243, fax +39 0903977464

Keywords: Acoustic emission, concrete, damage, Kohonen map, principal component analysis

Abstract

AE technique gained increasing interest in the last decade as a monitoring methodology and as an assessment tool for safety and reliability evaluation of reinforced concrete structures, historic and masonry buildings.

There are several established statistical methods (z, RA, b and Ib value) which can be used in analysing AE data with the aim to evaluate damage status of a structure subjected to a particular loading condition. Artificial neural networks (ANN) has recently applied as a tool to reduce data redundancy and to optimize feature set of AE signals.

In this paper cluster analysis to separate set of parameters into several classes reflecting internal condition was adopted to analyse acoustic emission data obtained during 4-points bending tests on concrete beams, under cycling and constant load condition, at increasing loads.

Two kinds of unsupervised clustering methods were used: the principal component analysis (PCA) and the self organized map (Kohonen map). Combining the results of these techniques, it has been possible to quantify the damage severity occurred to study the evolution of the damage itself during the test.

Introduction

The straight correlation between the propagation of acoustic waves due to cracking phenomena inside a concrete structure and the seismic wave propagation during an earthquake activity is a well known concept. Based on these similarities, starting from seismology studies, some statistical techniques with the aim to evaluate the “health condition” of concrete structure in a particular load condition have been developed. The \( b \) value analysis, from Gutenberg and Richter formula, originally outlines the relation between the magnitude of earthquakes with the number of events connected to them. In an AE structure monitoring test contest, this method and its improved evolution (Ib value) becomes a powerful tool for fracture modes identification. This kind of investigation doesn’t require the identification of sources location [1].

Seismicity rate changes have been used in great number of studies as a significant tool in order to explain the stress distribution in a specific area of the Earth’s crust. Some studies on precursors of past earthquakes suggest that particular space-time seismicity patterns, including the phenomenon of precursory quiescence, can be related to the seismo-tectonics process that lead to earthquakes. The quiescence hypothesis, as formulated by Wiss and Haberman [2], postulates that the quiet volume overlaps the main shock source volume. The seismic quiescence hypothesis assumes that some main shocks are preceded by seismic quiescence, that is a significant decrease of the mean seismicity rate (number of events of magnitude exceeding a given threshold, per unit time), as compared to the preceding background rate in the same crustal volume. This consideration has been transferred to cracking phenomena in concrete structure comparing the acoustic rate within a temporal window with a background rate [3].
As for fracture mode evaluation is concerning, different approach can be used, the most common one is based on the combination of two particular AE parameters (RA value and average frequency). [4]

Trying to define a common rule to identify the different damage mechanisms, particular AE descriptors like amplitude or duration can indeed be used. Sometimes, the analysis of a single parameter, or a combination of a few of them, is sufficient; but sometimes this procedure doesn’t provide immediately clear and coherent results. As a consequence, a discrimination of the acoustic signatures based only on one parameter is debatable, and a multi-variate technique appears an essential instrument for analysing data not immediately linked each others.

In the following sections the results of AE monitoring during concrete beam loading test will be described. Classical methodologies like $I_b$, $Z$ and RA values were performed to identify damage occurrence during the different loading tests. Furthermore, AE clustering methodologies like Principal Component Analysis and Self-Organize Map were also applied. Aim of this work is to develop a methodology investigation based on the use of different unsupervised algorithms that allow not only to analyse great amount of data to obtain a generic information about the structure monitored, but also, in synergy with other classical methodologies, allow to extract specific consideration about structural integrity of the tested specimen.

Experimental

![Four point bending test set-up](image)

A concrete beam of 15 MPa flexural strength was tested in a four point bending test (Figure1). The beam had a length of 500 mm and a cross section of 140x140 mm.

AE signals were recorded by a ten-channel Vallen AMSY-5 measurement system. A total 8 piezoelectric transducers for concrete (4 for each lateral side), VS30-V type with a flat response between 23-80 kHz were used. Threshold value after pencil lead break calibration was set at 40 dB: that allowed to eliminate the background noise and to record only the emissions due to cracking of the concrete. 4 main loading cycles were performed with increasing loads. Every cycle, moreover, was set up by 3 different load conditions (Table 1).

Every main cycle load starts with a 10 minutes rate loading condition and 10 minutes rate unloading condition for two reasons: Firstly we wanted to pre-crack the structure before applying a long time loading and secondly we wanted to evaluate the differences due to energetic release during unloading step compared with energetic store during loading step. Every cycle of the cycle loading was characterized by a frequency of 0.1 Hz. Constant load condition at the maximum load had a duration of about 300 minutes.
Table 1. Load history cycles

<table>
<thead>
<tr>
<th>TEST N°</th>
<th>MAXIMUM LOAD (kN)</th>
<th>LOAD CONDITION</th>
<th>DURATION (min)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>15</td>
<td>slow loading and unloading</td>
<td>21</td>
</tr>
<tr>
<td>2</td>
<td>15</td>
<td>cycle loading</td>
<td>340</td>
</tr>
<tr>
<td>3</td>
<td>15</td>
<td>constant loading</td>
<td>320</td>
</tr>
<tr>
<td>4</td>
<td>18</td>
<td>slow loading and unloading</td>
<td>21</td>
</tr>
<tr>
<td>5</td>
<td>18</td>
<td>cycle loading</td>
<td>390</td>
</tr>
<tr>
<td>6</td>
<td>18</td>
<td>constant loading</td>
<td>320</td>
</tr>
<tr>
<td>7</td>
<td>19</td>
<td>slow loading and unloading</td>
<td>21</td>
</tr>
<tr>
<td>8</td>
<td>19</td>
<td>cycle loading</td>
<td>370</td>
</tr>
<tr>
<td>9</td>
<td>19</td>
<td>constant loading</td>
<td>230</td>
</tr>
<tr>
<td>10</td>
<td>20</td>
<td>slow loading and unloading</td>
<td>21</td>
</tr>
<tr>
<td>11</td>
<td>20</td>
<td>cycle loading</td>
<td>410</td>
</tr>
<tr>
<td>12</td>
<td>20</td>
<td>constant loading</td>
<td>283</td>
</tr>
</tbody>
</table>

Ib, Z and RA values analysis

Adapting Gutenberg and Richter formula, in AE structure monitoring contest, it has been possible to distinguish the different fracture phases calculating the slope (b value) of the function:

\[ \log N = a - b A_{dB}/20 \quad \text{Eq. 1} \]

Where \( N \) is the number of AE events with an amplitude higher than \( A_{dB} \). The b-value analysis was applied to groups of subsequent events. In this work we decided to consider groups of 300 events detected by each sensor and divided for each load condition.

Implementing the formula with some statistical values of amplitude distribution like mean and standard deviation an improved b-value (Ib-value) was proposed by Shiotani [5]:

\[ Ib = \frac{\log_{10} N(w_1) - \log_{10} N(w_2)}{(\alpha_1 + \alpha_2)\sigma} \quad \text{Eq. 2} \]

Where \( N(w_1) \) is the accumulated number of AE events in which the amplitude is more than \( \mu - \alpha_1 \sigma \), and \( N(w_2) \) is the accumulated number of AE events in which the amplitude is more than \( \mu + \alpha_2 \sigma \). \( \sigma \) is the standard deviation of the magnitude distribution of one group of events, \( \mu \) is the mean value of the magnitude distribution of the same group of events, \( \alpha_1 \) and \( \alpha_2 \) are constants.

A parameter to describe changes in seismicity activity is the seismic quiescence. One commonly used test for comparing seismicity rate changes is Z-value test for a difference between two means. The Z-value measures the significance of the difference between the mean seismicity rates \( \mu_1 \) and \( \mu_2 \) within two time interval.

\[ Z = \frac{\mu_1 - \mu_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \quad \text{Eq. 3} \]

Where \( \sigma \) and \( n \) are, respectively, the standard deviation of the rate and the number of events in the group. In particular, the Long term average (LTA) is the Z value resulting from the comparison of the rate within a window, \( \mu_{Win} \) and the background rate, \( \mu_{All} \) defined as the overall mean rate in the volume. Negative Z values indicate rate increases while a positive Z corresponds to rate decrease. In this work we decided to consider groups of events detected in a temporal window of 2000 seconds by each sensor and divided for each load condition.

During a structure damaging, it is well known that the cracking fracture mode evolves from mode I type (nucleation and growth of tensile crack) to mode II type (shear propagation of crack) with the progress of fracture. RA value and average frequency are classical instrumental to classify active cracks. From AE parameters of the rise time and the peak amplitude,
The average frequency AF of a waveform is calculated as:

\[ \text{AF} = \frac{\text{Counts}}{\text{Duration}} \quad \text{Eq. 5} \]

**Clustering methodology of AE signals**

AE signals are generally described by 12 patterns, in particular, Amplitude, Counts, Duration, Risetime, Energy, are calculated by the acquisition system, while Historic Index, Severity, RA value and Average Frequency are parameters derived from previous ones. Time, Sensor and Load cycle were also other variables that were considered. The subsequent calculation of correlation coefficients implies the assumption that all AE features exhibit Gaussian distribution. To allow data comparison with different scales, logarithmic values were used instead of their natural values.

The Principal Component Analysis (PCA) is a factor model in which the factors are based on a reduced correlation matrix [8]. The data have been organized in a data matrix, addressing the problem of analyzing the structure of the interrelationships (correlation) among the large number of variables by defining a set of common underlying dimensions (factors). As a consequence, it was possible to identify the separate dimensions of the structure and then determine the extent to which each variable is explained by each dimension. We summarized the data (the data are described in a much smaller number of concepts than the original individual variables) and reduced the variables (calculating scores for each underlying dimension and substituting them for the original variables). The criterion used was connected to the evaluation of variance percentage. In our case 67% of total variance was evaluated and this led to realize PCA graphs using only three factors.

**Self-organizing neural networks (SOM)**

The self-organizing map is one of the most prominent artificial neural network models adhering to the unsupervised learning paradigm[9]. The model consists of a number of neural processing elements, i.e.units. Each of the units \( i \) is assigned an n-dimensional weight vector \( m_i \), \( m_i \in \mathbb{R}^m \).

The training process of self-organizing map may be described in terms of input pattern presentation and weight vector adaptation. Each training iteration \( t \) starts with the random selection of one input pattern \( x(t) \). This input pattern is presented to the self-organizing map and each unit determines its activation. Euclidean distance between the weight vector and the input pattern was used to calculate a unit’s activation. In this particular case, the unit with the lowest activation is referred to as the winner, \( c \), of the training iteration, as given in Equation 6.

\[ c: m_c(t) = \min_i \| x(t) - m_i(t) \| \quad \text{Eq. 6} \]

Subsequently, the weight vector of the winner as well as the weight vectors of selected units in the vicinity of the winner are adapted. This adaptation is implemented as a gradual reduction of the difference between corresponding components of the input pattern and the weight vector, as shown in Equation 7.

\[ m_i(t+1) = m_i(t) + \alpha(t) \cdot h_c(t) \cdot [x(t) - m_i(t)] \quad \text{Eq. 7} \]

The results of the training process is the so called U-matrix. This map shows distances from maps units and the nearest neighbourhood, evaluated by Euclidean method. High values of U matrix map (red ad yellow pixels) shows mean large distances between neighbouring map units.
elements belonging to the same cluster are identified by uniform areas of low value (blue pixels). In this specific case, we identified 3 main cluster areas (see the following figure 5a).

Additional information can be obtained plotting the hit Histogram U-matrix. This map shows the projection of data samples into map. The data are calculated by finding the BMU (Best Matching Unit) of each data sample from the map, and increasing a counter in the map unit each time it is the BMU. The colours are related to a specific level of a variable and hexagon size is related to the numbers of AE hits related to that cluster point.

Results

![Fig. 2. Ib and Z values vs typology load test](image)

In Figure 2 Ib value was characterized by a floating trend during loading test. Before the 8th load test, the curve has very low values. A crescent trend can be identified for each group of test characterised by same maximum load. As expressed in literature, the Ib low value corresponds to the macrocrack opening. Once the generated macrocracks opened up, most of the energy has been released. It implies the creation of many weak events leading to an increase of Ib value. In our results, during cycling loading tests low values of Ib are obtained, as a consequence of the activation of macrocracks. In the constant load tests the Ib value increases due to reduced energy released for crack formation and propagation. It could be related with micro-cracks events. The constant load conditions (3rd and 6th load tests) influence the trend of Ib value more than the cycling ones (2nd and 5th load tests). With the 8th load test (cycling load test), it’s possible to observe the generation of new microcracks (evidenced by relevant increase of Ib) that propagate and became macrocracks during the constant load condition (9th load test). This behaviour clarify that 19 KN group tests critically compromise the mechanical integrity of the concrete structure.

The Z and Ib statistics show a similar pattern. A significantly high Z-value is found at the 8th load test. Paraphrasing the seismologic terminology, this particular load condition could be considered like the beginning of a quiescent period preceding a severe damage condition of the structure.

The Z and Ib statistics show a similar pattern. A significantly high Z-value is found at the 8th load test. Paraphrasing the seismologic terminology, this particular load condition could be considered like the beginning of a quiescent period preceding a severe damage condition of the structure.

Figure 3 shows relations between RA value and average frequency of concrete specimen under all loadings. The data are clustered with a centroid representative of all data for all sensors and then classified for load cycle typology. Analysing the figure, it’s possible to observe two different behaviours: before the 8th load test, the plotted data are distributed in the area where the average frequencies are medium/high and RA values are low. This region is representative of mode I crack behaviour. After the 9th load test, AE events are observed in the region where the average frequencies are low. This region is typically connected with mode II crack behaviour. So, 8th and 9th load tests represent the transition from tensile type to shear type of fracture mode.
In Figure 4 the PCA plots are reported. In the graph, the vertical axis represents the unrotated factor II (component 2), and the horizontal axis represents the unrotated factor I (component 1). The axes are labelled with 0 at the origin and extend outward to ±0.5. The numbers on the axes represent the factor loadings. In this graph, all the variables are labelled. From visual inspection, it appears obvious the presence of 2 main groups: the variables Test, Load and Time (variables that we can generically define as time dependent) are grouped together and load fairly high. The variables strictly connected with the wave properties (Risetime, Duration, Counts and Energy) generates a second group very high on the unrotated second factor. Moreover, analysing the particular data distribution, we can consider them as organized in the three clusters strictly connected with time dependent variables trend. This outlines how strong is contribution of such variables as data discriminating factors. Irregular are the distribution of other variables like...
Amplitude, RA, Historical Index, AF and Severity. Analysing the third graph, on the unrotated third factor, Amplitude loads high. A little more difficult is the comprehension of groups analysing the second graph.

For a better comprehension of these results, factor loadings greater than ± 0.3 are considered to meet the minimal level of significance; loadings greater than ± 0.4 are considered practically significant. Comparing PCA analysis results with variables correlation matrix, Load and Test variables can be removed, because correlated with Time variable, and also, Risetime, Counts and Energy can be removed, because equivalent to Duration. Using this residual variables and the ungrouped ones, the SOM was realized.

![Fig. 5 a) U-matrix resulting from application of the SOM; b) Hits U-matrix for the variable Load; c) Hits U-matrix considering as variables the cycling load conditions](image)

In Figure 5b the hit Histogram U-matrix is related to the variable load. Cluster 1 is mainly related to 19 kN load condition (blue hexagons) and as discussed before, this region represents the area in which weak events increase \( I_b \) value as a consequence of microcrack development, and considering Z value results, this area is characterized by a revelation of a quiescence period. Cluster 2 is related to 15 kN load condition (red hexagons) and 18 kN (green hexagons) and this area is connected with nucleation and growth of macrocracks (mode I fracture behaviour). Cluster 3 is related to 20 kN load condition (yellow hexagons) and as reported before, this represents area in which existing cracks are propagating in shearing mode (mode II or III fracture behaviour).

In Figure 5c is represented the hits U-matrix related to specified test conditions. As before mentioned, the 18 kN cycling load condition, identified as 8\(^{th}\) load cycle, outlines the transition from tensile type to shear type of fracture mode. Trough the Hits U-matrix, it was possible to localize inside the map hits connected to this particular load condition and compare this positions with the hits representative of two other cycling load conditions. In particular we reported the hits connected to 18kN cycling load condition (5\(^{th}\) load cycle) and 20 kN cycling load condition (11\(^{th}\) load cycle). As it is possible to observe, the 8\(^{th}\) load cycle (red hexagons), is mainly present in Cluster 1 region and, as mentioned before, we can identify this area, as quiescent seismic region; the 5\(^{th}\) load cycle (blue hexagons), is present in Cluster 2 region (mode I cracks area) while the 11\(^{th}\) load cycle (green hexagons) is mainly present in Cluster 3 region (mode II cracks area).
Interesting considerations can be done comparing U-matrix with topological map of single variables (Figure 6). The cluster regions evidenced in U matrix are quite similar to Time Map, outlining the strong weight of this variable compared with others. Similar considerations came out in PCA analysis about time dependent variables. Sensor Map is characterized by homogeneous distribution of results, this correspondence confirms the high coherence in data acquisition by each sensor during all loading tests. Furthermore, analysing Severity Map, the highest severity conditions are connected with Cluster 3 region. In this region we identified hits connected to 20 kN load condition and, referring to Z value results, in this area, exceeding quiescence condition, the structural integrity of specimen starts to be compromised. Minimum values corresponds to data inside Cluster 1 region (intermediate test condition loadings).

Conclusions

Clustering algorithms were used and compared with traditional procedures of damage evaluation (Ib, Z and RA Values). Principal Components Analysis allowed to correlate clusters to groups of variables and to identify related variables. In such a way a reduction of original data matrix was obtained. The Kohonen’s self organizing map algorithm was adopted to identify clusters and relate them to the AE patterns. Although the use of these new methodologies appears promising as a practical tool of evaluation of many different parameters in a restricted number of graphs , it requires a research of a valid procedure to optimize a correct interpretation of great amount of data, even requiring integration with classical well established procedures.

References

2. Wyss M, Habermann R. E.: Precursory seismic quiescence, 1988; 126, 319-332
5. Shiotani T., Yuyuma S., Li Z., Ohtsu M.: Journal of Acoustic Emission, 18, 118-133