Improved signal processing of acoustic emission for structural health monitoring using a data-driven approach

Sebastian F Wirtz and Dirk Söffker
Chair of Dynamics and Control, University of Duisburg Essen, Germany, sebastian.wirtz@uni-due.de

Abstract

Extracting isolated signal patterns from continuously sampled data and discarding noise is of particular importance in the field of Structural Health Monitoring. Using Acoustic Emission (AE) for passive in-situ inspection, this is considered as a challenging task. Here, ultrasound stress waves are recorded passively to detect damage occurring in a material or structure in real-time. The corresponding waveforms recorded by AE transducers carry information regarding their underlying source mechanism, which can be related to either damage or spurious noises from the environment. Typically, these signals are transient and stochastically distributed in time. Many state-of-the-art AE systems use fixed thresholds to distinguish between relevant signals and noise. However, in practical applications these approaches are inaccurate or even unfeasible due to low signal-to-noise ratio. In this paper, a data driven approach to distinguish between relevant signals and noise is presented. To identify a model from the AE data Platt calibration is used, which was developed to map the output of support vector machines to probabilities. The advantage of this approach is that the mapping between signal characteristics and relevance can be derived from preliminary measurements such as Pencil-Lead Break (PLB) test. In a first step, the distance of the measurements to a baseline is computed. Furthermore, by determining the onset of the PLB signal manually or using a different method such as Akaike information criterion, the measurement can be split in two disjoint sections containing the relevant signal and noise, respectively. Finally, a parametric model is derived from the obtained class-conditional distributions of distances using a curve fitting procedure. Based on this model a characteristic function which serves as a probabilistic measure indicating the presence of relevant signals can be computed. The feasibility of the approach is demonstrated using AE measurements from indentation tests of composite plates.

1. Introduction

Acoustic Emission (AE) testing is a passive, in-situ inspection method which originates from the field of Non-destructive testing (NDT). Here, ultrasound stress waves are recorded passively to detect damage occurring in a material or structure. The corresponding waveforms carry information regarding their underlying source mechanism, which can be either related to damage or spurious noises from the environment. Typically, these signals are transient and stochastically distributed in time.

In contrast to NDT applications, the main focus of Structural Health Monitoring (SHM) is the continuous monitoring of structural components (1). In this field, AE is particularly suitable due to high sensitivity to local, incipient damages and because it can be ap-

Creative Commons CC-BY-NC licence https://creativecommons.org/licenses/by-nc/4.0/
plied in-situ. Besides SHM, the interest in continuous AE measurements is also growing in related fields of research. For example in material science continuous AE monitoring of fatigue experiments can provide valuable information regarding underlying damage mechanisms which might be unavailable in case of intermittent measurements (2). However, because large amounts of data are acquired during continuous AE measurements within a short amount of time, the ability to discard noise can reduce storage requirements and processing time. This is particularly important with respect to waveform-based approaches, where increasingly complex algorithms for data analysis including time-frequency domain transformation and multivariate statistical methods are used. In contrast, many state-of-the-art AE systems use fixed thresholds to distinguish between relevant signals and noise. However, in practical applications this approach is inaccurate or even unfeasible due to low signal-to-noise ratio.

The main contribution of this paper is the development of a method-oriented but practical methodology which can be used to determine a suitable threshold to distinguish between relevant subsections of AE measurements and noise. Here, an implicit definition of the decision threshold allows taking into account data obtained from preliminary measurements such as Pencil Lead Break (PLB) test. Furthermore, it has to be noted that in contrast to the related field of waveform picking, the focus is to determine a relevant subsection and not the exact onset time. The rest of the paper is structured as follows. Background and related work are presented in section 2. The procedure of the data-driven methodology is illustrated in section 3 and results of experimental evaluation are given in section 4. Finally, summary and conclusions are given in section 5.

2. Background and related work

In general AE testing relies on passive monitoring of transient signals, which are related to elastic stress waves from different source mechanisms. The related signals are typically characterized by high frequency content and low amplitudes. Furthermore, relevant signals are transient and stochastically distributed in time. Therefore, large sections of noise are typically obtained in case of continuous monitoring. Research regarding signal processing of AE can be divided into (i) source characterization and (ii) waveform picking. In the following, a brief overview of the related works is presented.

To distinguish between different source mechanisms, features are used to emphasize characteristic properties of the related signals. By extracting suitable features, the interpretability of AE measurements can be enhanced. A detailed summary is given by Kaphle et. al. (3). In time domain heuristic parameters are typically determined from the measurements with respect to a fixed threshold. These include for instance maximum amplitude, rise-time, or energy. In contrast, waveform-based analysis refers to techniques, which consider the recorded waveforms as a whole. Here, signal features can be computed directly from raw data (e.g. higher order statistical moments) or using transformed representations in different domains (e.g. frequency- and time-frequency domain). Additionally, new parameters such as entropy and Chebychev moments were proposed for analysis of AE waveforms recently (4,5).
In contrast, waveform picking focuses on finding the exact onset time of an AE waveform. Waveform picking is closely related to the task of AE source localization, where time-of-flight measurements are used to relate individual AE signals to their points of origin in a structure. Here, a simple and in the field of AE testing widely adopted approach is amplitude thresholding. However, this approach is likely to provide inaccurate results leading to errors in the source location.

Many advanced waveform picking approaches have their origins in the field of computational seismology (6). Typically, a characteristic function is defined, which is used to emphasize the change of statistical properties of the measurements due to the presence of a signal. Then a user-defined threshold can be applied using the characteristic function instead of the raw data. For example, using the Short Term Average (STA) / Long Term Average (LTA) method, a characteristic function is defined using the ratio of average amplitudes within windows of different length (6). Thus, instantaneous changes in the amplitude distribution lead to peak values in the characteristic function. The onset of the AE signal can be detected by choosing a suitable threshold. Similarly, Rawles et al. define a new characteristic function, which is based on the direct comparison between raw data and a reference set using a shape-based similarity measure (7). Furthermore, Pomponi et al. proposed a methodology based on wavelet block thresholding for onset detection (8). Here, a thresholding rule is applied to individual subsections of each decomposition level leading to a binary mask that defines which coefficients are retained. Also, the binary mask can be used to detect the onset of signals by finding the first non-zero element (8). In contrast, using Akaike Information Criterion (AIC) choosing user-defined thresholds is not required (6). Here, the onset of the signal is determined as the global minimum of a characteristic function that can be computed sample-wise from raw data. Also, Bai et al. developed an alternative approach which is independent of user-defined thresholds based on surrogate significance test (9). Here, the signal onset is determined as the intercept of different correlation functions.

3. Methodology

In this section, a technique to extract relevant subsections from AE measurements is presented, which utilizes example data to determine a suitable detection threshold. As illustrated in Figure 1, the observed amplitudes of a continuously sampled AE signal belong to different probability distributions which are related to either relevant AE signals (i.e. transient waveforms referred to as hit) with superimposed noise or plain noise. However, the individual distributions are strongly dependent on the experimental conditions and the environment. Therefore, the choice of a suitable threshold is a trade-off between the sensitivity of the measurement system and rejection of noise and usually based on the experience of the operator. The underlying idea of the proposed approach is to estimate the probability that a given sample belongs to an AE signal using a parametric model. The parameters of the model can be identified from suitably chosen example data. Finally, the obtained model can be used to calculate a characteristic function to distinguish between relevant signals and noise by choosing a probability threshold. Thus, the related threshold in terms of the decision variable is defined implicitly, because it depends on the model parameters which are determined according to the example data.
To identify a model from the AE data Platt calibration is used, which was developed to establish a mapping between numerical scores computed by support vector machines and probability estimates of the related classes (10). As illustrated in Figure 1, two overlapping distributions of the decision variable are assumed. As a model the sigmoid function

\[ f(x) = \frac{1}{1 + e^{Ax + B}} \]  

is used, where the parameters A and B are determined by a curve fitting procedure.

In the following, the AE measurement \( S_{1:n} \) is assumed as a sequence of samples \( s_i \) in discrete-time

\[ S_{1:n} = \{ s_i, i = 1, 2, \ldots, n \} \]  

where the duration between each sample is determined by a fixed sample rate. Each sample can belong to one of two classes i.e. noise \( (y = 0) \) or presence of AE signal \( (y = 1) \). In the sequel, signal probability refers to the likelihood that a given sample \( s_i \) belongs to an AE waveform. As decision variable the similarity of the measurement to a noise sample (baseline) is considered. For a given signal, the city block distance

\[ d(S_{1:n}, N_{1:n}) = \sum_{i=1}^{n} |s_i - n_i| \]  

is computed with respect to the noise sample using sliding window.

To obtain an estimate of the class conditional probability \( p(d \mid y) \), conditional densities \( P(y = \{0, 1\} \mid d) \) are determined after splitting the dataset into sections containing noise and relevant signals according to AIC criterion using Maedas formula (6)

\[ \text{AIC}(i) = s_i \log(\text{var}(S(1, s_i))) + (N - s_i - 1)\log(\text{var}(S(s_i + 1, n))) \]  

Finally, the parameters A and B of the sigmoid function are determined using least squares method. Given the parameters A and B, an estimate of the signal probability can be calculated as characteristic function by evaluating (1) sample-wise. Relevant subsec-

\[ \text{Figure 1. Illustration of the class-conditional amplitude distributions and mapping to signal probability.} \]
tions $S_{a:b}$ of AE measurements can be extracted by searching for contiguous blocks of samples exceeding a signal probability threshold $T$ as

$$S_{a:b} = \{ s_i \in S_{1:n} : s_i > T \}.$$  

In this case, the decision threshold is defined implicitly in terms of signal probability. The related threshold applied to the decision variable depends on the parameters $A$ and $B$ and thus is adjusted depending on the provided preliminary measurements.

4. Experimental evaluation

In this section, results of the experimental evaluation are presented. As an example, AE data from indentation testing of composite material is used. To detect relevant AE signals, the parameters of the related model are determined from PLB tests. The sensitivity of the new approach is compared to an energy-based approach. Finally, examples of extracted waveforms are presented.

4.1 Experimental setup

The data which are used in the following are obtained from indentation testing experiments as described by Baccar and Söffker (11). The mechanical test rig is illustrated in Figure 2. Here, a plate with the dimensions 425 mm x 425 mm x 1.8 mm of laminated cross-ply composite material with woven fabric on the outer layer is fixed horizontally below the indentation tool by using a custom clamping system. The AE sensor is stiffly bonded to the surface of the composite plate using cyanoacrylic glue. During the experiments, load is applied manually in transverse direction.

![Figure 2. Illustration of the experimental setup (11).](image)

In this work, a sharp, cone-shaped tool was used during the indentation test, which penetrates the material on increasing load. An example of a specimen after the test is shown in Figure 3. On the top side of the plate (Figure 3 (a)) typical indentation marks are visible at the surface. Apart from that, the surface remains undamaged. In contrast, extended damage can be observed on the bottom side (Figure 3 (b)). Here, buckling of the surface in a larger area indicates extensive delamination and damage to the matrix material. Furthermore, broken fibers of the woven fabric and a crack extending towards the top of the figure are visible.
4.2 Model fitting results

To determine the parametric model, measurements from three independent PLB tests are used. In Figure 4, results of fitting the model to the experimental data are illustrated. Figure 4 (a) shows the obtained class conditional distributions of the decision variable after splitting the measurements into noise and relevant signals (hit) using AIC according to equation (4). Both distributions are skewed to the right and an overlap can be observed. However, the means are clearly separated. Compared to the amplitude distribution of relevant signals, the distribution related to plain noise is narrow and located at smaller values of the decision variable. The corresponding sigmoid function, which is used to model the mapping between the decision variable to signal probability is presented in Figure 4 (b).

Furthermore, results of several independent runs are summarized in Table 1. Here, additional artificial noise is added to the original measurements to investigate the effect of different SNR. It is noticeable that the absolute value of the parameter A, which controls the gradient of the sigmoid function, decreases due to lower SNR. Accordingly, the resulting sigmoid is stretched so that the corresponding signal probabilities are shifted towards larger values of the decision variable. Thus, sensitivity of the characteristic function is decreased in favor of improved noise rejection. Therefore it can be concluded that the adjustments of the model parameters with respect to the example data are reasonable.
### Table 1. Parameters determined for three independent PLB

<table>
<thead>
<tr>
<th>SNR [dB]</th>
<th>A</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td>60</td>
<td>-3191.80</td>
<td>7.13</td>
</tr>
<tr>
<td>30</td>
<td>-1512.82</td>
<td>9.69</td>
</tr>
<tr>
<td>25</td>
<td>-605.87</td>
<td>6.12</td>
</tr>
</tbody>
</table>

#### 4.3 Sensitivity analysis

Given the model parameters, the signal probability can be estimated by evaluating the sigmoid function sample-wise. The result can be used as a characteristic function to detect relevant subsections of the AE measurement. In the sequel, the sensitivity of the signal probability as a characteristic function is compared to the energy envelope using the Receiver Operator Curve (ROC). The ROC is a method which originates from field of detection theory and is frequently used to assess the sensitivity of classification algorithms by comparing the behavior of true positive and false positive rates depending on the threshold (12). However, for this type of analysis a ground truth is required.

In Figure 5, a subsection of a measurement obtained during early stage of indentation test prior to the fracture of the specimen is shown. To obtain a ground truth for ROC analysis, this section has been labelled manually. Samples which were labelled as relevant AE signals are highlighted in green color.

![Figure 5. Ground truth for ROC evaluation.](image)

In Figure 6 the results of the ROC analysis are presented for different values of SNR. The top right corner of the ROC corresponds to the lowest threshold, where each sample is detected as a relevant signal. Moving along the curve to the left, the ROC shows the behavior of true positive and false positive rates on increasing thresholds. Here, a decrease in false positive rate can only be achieved at the cost of a reduction in true positive rate. Improved performance of the detection algorithm is achieved as the ROC approaches the top left corner (i.e. high true positive rates are maintained while reducing the false positive rate). Therefore, the area under the ROC is frequently used to assess sensitivity of classification algorithms (12). According to Figure 6, both estimated signal probability and energy envelope show decreasing areas under the ROC in case of lower SNR, thus indicating a reduction in sensitivity due to the effect of noise. However, especially in the range of lower thresholds the ROC of the estimated signal probabil-
ity as characteristic function (solid line) is clearly located above the ROC obtained using energy envelope (dashed line) indicating improved detection rate. Also, regarding the lowest SNR, the solid line indicates improved detection rate of the estimated probability over a wide range of possible thresholds. Thus, it can be concluded that in comparison to the energy envelope improved sensitivity can be achieved using the estimated signal probability.

![ROC curves](image.png)

**Figure 6. Comparison ROC curves using signal probability and energy envelope for different SNR.**

Moreover, extraction of relevant waveforms from an AE measurement can be realized by searching for contiguous subsections of the signal exceeding a probability threshold. Using equation (5), relevant subsections were extracted from the subsection presented in Figure 5. Two examples of AE waveforms obtained at SNR of 30 dB and threshold of 0.1 are presented in Figure 7. In the first example (Figure 7 (a)) the signal amplitude is large compared to the noise floor leading to a corresponding mean signal probability of 80.8 %. In contrast, the distance between the signal and the noise floor is smaller. Accordingly, lower mean signal probability of 36.27 % is obtained (Figure 7 (b)).

![Extracted waveforms](image.png)

**Figure 7. Examples of extracted hits (a) high probability (b) low probability.**
5. Summary and conclusion

In this paper, a new methodology to extract relevant waveforms from AE measurements using Platt calibration is presented. Here, a parametric model of the data is determined using preliminary measurements. To distinguish between noise and relevant signals, the detection threshold is defined implicitly in terms of signal probability and adjusted according to the provided data. Experimental evaluation of the approach is presented at the example of AE measured during indentation test using laminated composite material. The results of multiple runs using different SNR indicate that the parameters of the fitted model are adjusted reasonably depending on the SNR. Furthermore, the sensitivity of the methodology is compared to the energy envelope of the AE signal using ROC. From the results it can be concluded that improved sensitivity can be achieved for low thresholds.

References

