TU Delft

Delft University of Technology

Document Version
Final published version

Licence
CC BY

Citation (APA)
Hohaus, L., Kassapoglou, C., & Pahlavan, L. (2026). Use of Principal Component Analysis and Autoencoder Neural Networks in Characterization of Acoustic Emission Waveforms from Composite Coupon Specimens. In *Proceedings of the 12th European Workshop on Structural Health Monitoring (EWSHM 2026): July 7-10, 2026 in Toulouse, France* Article 1505 (e-Journal of Nondestructive Testing; Vol. 2026, No. 08). NDT.net. https://doi.org/10.58286/33836

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright
In case the licence states “Dutch Copyright Act (Article 25fa)”, this publication was made available Green Open Access via the TU Delft Institutional Repository pursuant to Dutch Copyright Act (Article 25fa, the Taverne amendment). This provision does not affect copyright ownership.
Unless copyright is transferred by contract or statute, it remains with the copyright holder.

Sharing and reuse
Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy
Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

This work is downloaded from Delft University of Technology.

More info about this article: https://www.ndt.net/?id=33836

Use of Principal Component Analysis and Autoencoder Neural Networks in Characterization of Acoustic Emission Waveforms from Composite Coupon Specimens

Leonard HOHAUS[1], Christos KASSAPOGLOU[2], Lotfollah PAHLAVAN[1]
[1]Ship and Offshore Structures, Mechanical Engineering, TU Delft, Netherlands
[2]Aerospace Structures and Materials, Aerospace Engineering, TU Delft, Netherlands

Abstract Acoustic emission (AE) monitoring is a method of structural health monitoring that relies on the detection of elastic waves generated by the release of concentrated strain energy when damage is created in a structural material. One of its strengths is that recorded waveforms can, in theory, be used to infer the source mechanisms, enabling estimation of damage type, location, and severity. One of the challenges, however, is how to automate the interpretation and reliably retain useful information from AE waveforms, which are often thousands of samples long. In this work, a method of waveform analysis in the frequency domain is presented that combines principal component analysis and an autoencoder to reduce the dimensionality of the problem to a pair of parameters, capturing the spectrum shape and the spectrum energy content. Tensile tests were carried out on composite coupon specimens while monitoring AE, and the progression of damage was assessed by X-ray scanning the specimens at two locations along their length, both before and after testing. Locations and types of damage in the scans are in good agreement with the results of the AE monitoring and analysis framework. The method is proposed as a tool for automated interpretation of AE signals with the potential to be generalized to other material, layup, and sensor setups.

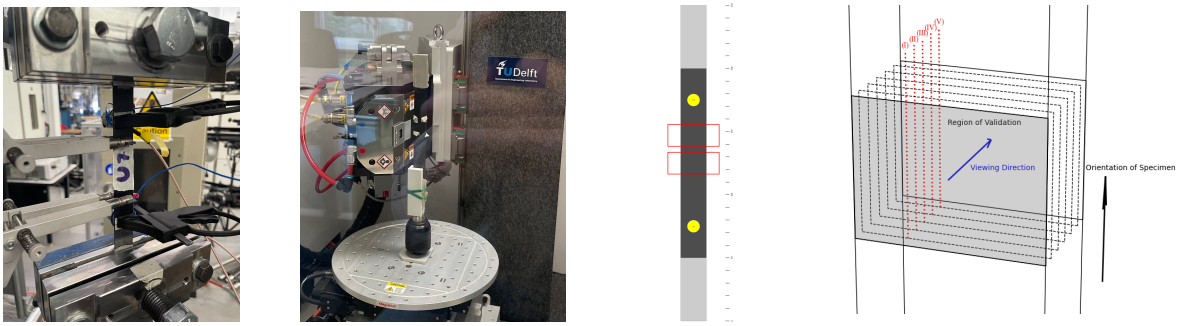
Keywords: Acoustic Emission; Waveform Analysis; Principal Component Analysis; Autoencoder; Damage Characterization; Localization; Micro-CT

Introduction

Characterization of acoustic emission (AE) data is a desirable feature of a robust AE-based structural health monitoring system [1, 2]. Especially so, when dealing with fibre-reinforced composites, which suffer from a variety of possible damage mechanisms, such as fibre fractures, matrix and interface cracks, and delaminations. They can range from very harmful to less severe for structural performance [3]. One big challenge in the interpretation of AE data is that labeled datasets reliably reporting waveform characteristics of different types of damage are very difficult to obtain, and the transferability of characteristic features across different materials, layups, and monitoring setups is very limited [4].
Recent advances in the field of unsupervised machine learning, however, have made this objective more attainable as improved and more efficient implementations of automated pattern recognition have emerged [5].
In this work, we propose a method of analysis of acoustic emission signals that uses normalized frequency spectra and spectrum energy content separately to characterize internal damage events. The idea of using normalized frequency spectra originates from a musical analogy (in which an AE sensor corresponds to a musical instrument and different damage mechanisms to different chords played on that instrument), where distinct notes should be

This work is licensed under CCBY 4.0

Media and Publishing Partner NDT.net
https://doi.org/10.58286/33836



(a) *U7 in tensile test machine.* **(b)** *U7 inside micro-CT scanner.* **(c)** *Scanned regions* **(d)** *Scan image orientation*

Fig. 1. *Experimental setup and schematics of scan locations and orientation.*

recognizable whether they are played forte or pianissimo, i.e. independently of their amplitude. The method is based on a combination of principal component analysis and autoencoder neural networks.
Tensile tests were carried out on two fibre-reinforced composite coupon specimens while monitoring acoustic emissions using surface-mounted piezoelectric sensors according to the procedure detailed in Section 1. Analysis and localization of AE events along the intra-sensor distance were performed using the methods detailed in Section 2, and micro-CT scans are shown in Section 3, which were carried out to compare the results of the AE signal analysis with the actual progression of internal damage.

1. Experimental Setup

1.1 Composite Coupon Specimens

Two ASTM D3039 [6] unidirectional [0°] carbon-fibre reinforced coupon specimens (*U7* and *U9*) were manufactured for tensile testing from a carbon-fibre-epoxy-resin-prepreg system for out-of-autoclave curing (TORAY Composites 2510 [7]). The dimensions deviated slightly from the norm, with a width of 20 mm rather than the specified 15 mm to ensure proper mounting of the piezoelectric AE sensors on the surface.
The specimens were CT-scanned after manufacturing and showed large defects and inclusions at the interfaces between the layers. This can be attributed to the manual hand layup process and the expired prepreg system used for manufacturing. Defects are generally seen as weak spots of probable premature damage growth [8], and, therefore, these inclusions were not seen as a cause of concern, as the objective was to cause damage growth in the first place.

1.2 Acoustic emission measurement system

AEs were monitored on the specimens during quasi-static tensile tests using two piezoelectric WSα (100-1000kHz) wide-band AE sensors by MISTRAS Group [9]. The transducers were mounted on the specimens using manual mechanical clamps (Fig. 1a), and silicone grease applied between the surface of the transducers and the specimens to create an acoustically coupled interface. 40dB pre-amplifiers connected sensors to the data acquisition, for which *Vallen Systeme*'s multi-channel AMSY-6 system was used. When recording a hit, stored information includes the sampled waveform, the time of arrival, and the detection threshold (25.3 dB).

1.3 Test and validation procedure

The specimens were tested in tension up to 10 kN, corresponding to a strain of $\sim$ 0.4 %, using a Zwick Universal Testing Machine to induce first-cycle micro-damage initiation while recording the AEs.

This work is licensed under CCBY 4.0

Media and Publishing Partner NDT.net

Micro-CT scans with a resolution of $8\,\mu m$ were performed before and after tensile testing to investigate internal structural changes caused by the loading and cross-check with the AE monitoring framework. The scanner used was the TESCAN CoreTOM Micro CT machine shown in Figure 1b.

2. Data Analysis Methods

The topology of the AE waveform dataspace is theoretically governed by the sensor characteristics and data acquisition settings, such as sampling frequency and filters. The distribution of data points within that space, however, depends on event frequency content, which in turn can be affected by event characteristics such as active AE source mechanism, path of transmission through the structure, and the interface between the structure and the sensor. In the following sections, we show how the dataset was constructed, and its topology explored using principal component analysis and an autoencoder neural network. AE events were thus located in that space using the autoencoder's single latent variable z as a proxy for source mechanism type. The influence of propagation channel (the specimen), AE transducer response functions, and transducer coupling variability on detection discrimination were disregarded in the current work on account of the small size of the monitored length ($100mm$) and the similarity (up to small variations) of setup conditions across the two loading experiments - the same sensors, type of coupling gel, sensor clamping force, and specimen boundary and loading conditions were used to perform AE measurements in both experiments.

2.1 Signal preprocessing and feature representation

The dataset of AE waveform signals ($[\mathbf{x_n}]$, sampling frequency $f_s = 2.5\,MHz$, fixed length $N = 2048$) was prepared for dimensionality reduction using the following sequence of signal processing methods: Discrete Fourier transform (Eq. 1a), amplitude normalization by energy content E (Eq. 1b) up to the Nyquist frequency index n_f, and centering by subtracting the mean normalized frequency spectrum $\hat{S}_{mean}$.

$$S_k = \sum_{n=0}^{N} x_n \, e^{-i2\pi \frac{k \cdot n}{N}}, \qquad k = 0, \ldots, N \tag{1a}$$

$$\hat{S}_k = \frac{2\,|S_k|}{\sqrt{E}}, \; E = \sum_{m=1}^{n_f} (2\,|S_m|)^2 \, \delta t, \; \delta t = \frac{1}{f_s}, \quad k = 1, \ldots, n_f, \quad n_f = \left\lceil \frac{N+1}{2} \right\rceil \tag{1b}$$

2.2 Principal component analysis for pattern recognition

Principal Component Analysis (PCA) is an unsupervised machine learning method that returns the linear principal components of a dataset, which are the orthogonal, i.e., non-redundant, directions of largest variance from the mean feature vector [10]. In this work, PCA has been applied to the dataset of normalized frequency spectra directly, without extracting parametric signal features in advance. Representing normalized AE waveform spectra in the space of the first few principal directions can 'spatially' separate qualitatively different subgroups of data and highlight their intrinsic structure using only a handful of coordinates [11].
The first six principal components obtained from the normalized dataset of specimen U7 are shown in Figure 2. The motivation for this choice is discussed in Section 4.

This work is licensed under CCBY 4.0

Media and Publishing Partner NDTnet

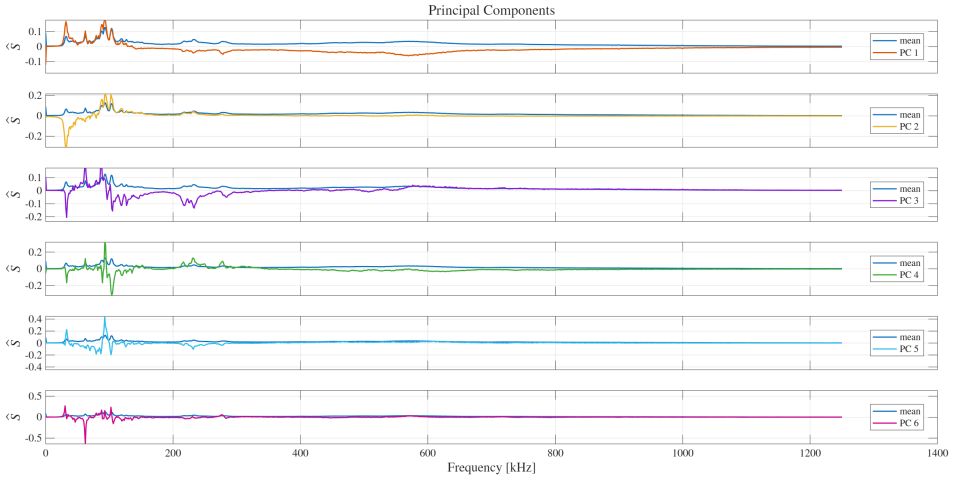
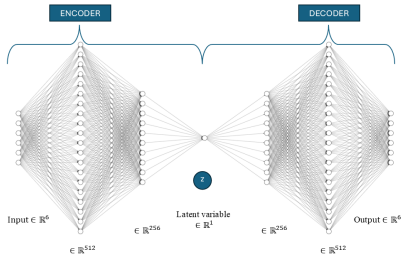


Fig. 2. *First six principal components of the normalized AE frequency spectra from specimen U7*

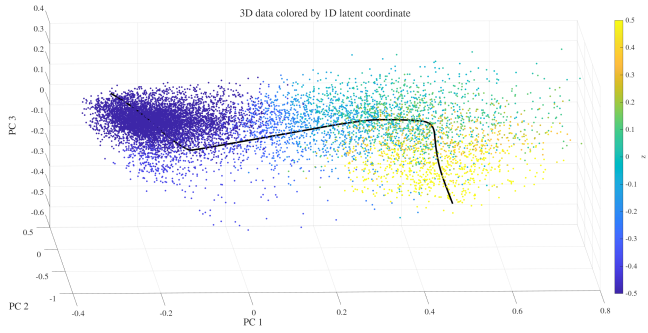
These six vectors were used as an orthonormal basis to reduce the dimensionality of the complete dataset of normalized frequency spectra from $n_f = 1025$ to $d_r = 6$ dimensions. The complete dataset in 3D principal component space is shown in Figure 3b.

2.3 Autoencoder neural network for parametrization of non-linear dataset manifold

Autoencoders are a class of neural networks used for non-linear data dimensionality reduction [12]. Their basic training architecture consists of a combination of an encoder and a (specular) decoder network, such that they can be trained to reconstruct their



(a) *Autoencoder architecture*



(b) *AE dataset in principal coordinates parametrized by latent variable z.*

Fig. 3. *Architecture and application of autoencoder*

input as accurately as possible (see Fig. 3a). The key is that there is a central hidden layer dividing encoder and decoder networks, which contains fewer nodes than the input and output layers. This forces the entire information to pass through a hidden bottleneck before it is reconstructed as accurately as possible at the output layer, and results in a non-linear mapping from the input space to the latent space, and its inverse [13].
In this work, an autoencoder with a single latent variable (shown in Figure 3a) was trained on the reduced datasets ($d_r = 6$) from both specimens, which included about 10 000 signals. Then, the encoder network, i.e. the first half of the autoencoder, was used to characterize the signals in terms of z (representing the signal's position in PC space). The non-linear curve traversing that empirical data space for $z \in [-1, 1]$ (black curve in Fig. 3b) was reconstructed using the decoder part of the neural network.

2.4 Physical source localization along the length of the specimen using TDOA

The time difference of arrival (TDOA) method was applied to locate AE sources along the length of the specimen. The group speed of flexural A0 waves in the unidirectional carbon fibre coupon specimens is nearly constant at $c_g \approx 2100\ m/s$ in a range of frequencies between 200 and 250 kHz, so waveforms were band-pass filtered in that range, and an envelope estimation was used to determine the true time of arrival of the ultrasonic wavepackets at each sensor. ΔT_i is the difference in time of arrival of an AE event i at the two sensors, and was used to estimate Δx_i, the source distance from the center of the specimen according to Eq. 2. The method was validated by repeated Pencil Lead Break tests on the surface at three known locations, which were identified with a mean square error of 2.25 mm on a measurement distance of 100 mm.

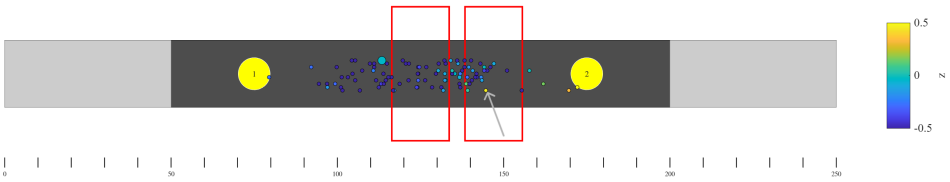
$$\Delta x_i = \frac{\Delta T_i \cdot c_g}{2} \tag{2}$$

The guided wave group velocities c_g were computed in advance using an in-house implementation of the Semi-Analytical Finite Element Method (SAFEM) [14, 15] tailored for a 1 mm thick unidirectional carbon-fibre composite with propagation along the direction of the fibres.

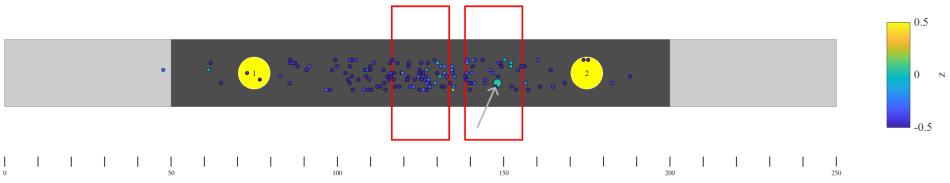
3. Results

3.1 AE source characterization and localization along the specimens

The results for AE events recorded by both sensors during loading of the two specimens are shown in Figure 4 using dots of different sizes and colors. Dot radius indicates logarithmic energy magnitude (parameter E), and dot color represents z, which describes spectrum shape type or 'character'.



(a) *Specimen U7*



(b) *Specimen U9*

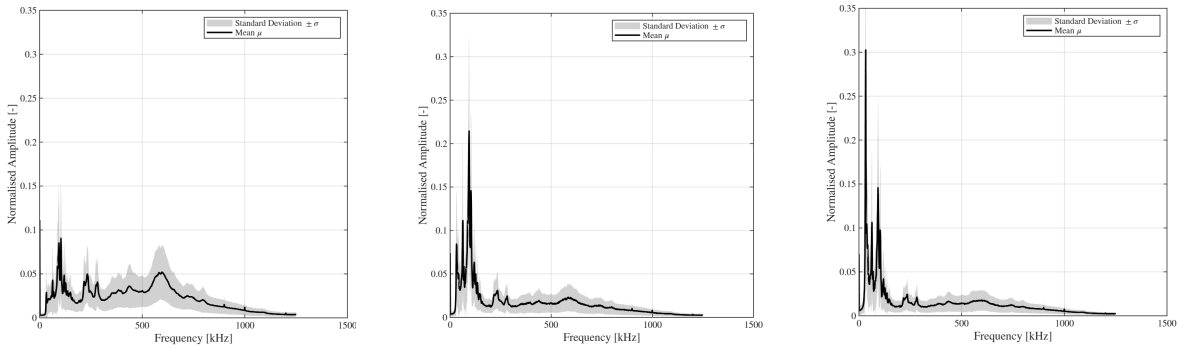
Fig. 4. *All AE events recorded during tensile loading tests characterized and localized along the length the two specimens. Gray arrows indicate events* ***(a)*** *and* ***(b)*** *from Section 3.1.*

The z-value is an indicator of signal frequency content and was observed to consistently characterize signals according to the following scheme:

- $z < 0$: Low signal-to-noise (SNR) ratio and high-frequency signals with spectral contributions beyond 500 kHz (Fig. 5a).
- $z \approx 0$: High-frequency head at 90 kHz followed by a low-frequency tail (Fig. 5b).
- $z > 0$: Low-frequency signals, dominant contribution at 30 kHz (Fig. 5c).

All normalized frequency spectra were clustered across these z-defined subsets. Mean and standard deviation of the three subsets are shown in Figure 5. One detail worth noting is

the three small yet visible spikes present in the mean normalized spectra of all three subsets at the frequencies of 900 kHz, 1000 kHz, and 1200 kHz. These peaks coincide with the cut-off frequencies of higher-order Lamb modes computed using dispersion relations from SAFEM [14] for the material and unidirectional layup of 1mm thickness used in this study. The authors suggest the specimen may experience a zero-group velocity (ZGV) Lamb mode at these frequencies, causing a sharp amplitude gain in the recorded signals, though further investigation via controlled experiment is required to confirm this.



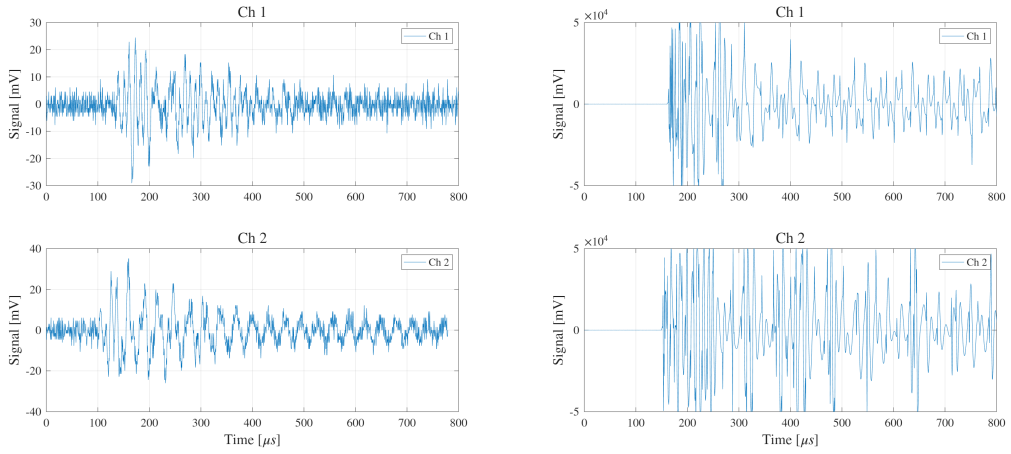
(a) $z < -0.05$ **(b)** $z \in [-0.05, 0.05]$ **(c)** $z > 0.05$

Fig. 5. *Average and variability of normalized frequency spectra grouped by z.*

Within the scanned regions, most registered AE events fell into Group 1, i.e. $z < -0.05$ (see Fig. 4), corresponding to small-magnitude events with low-SNR signals. Owing to the relatively large variability of frequency spectra within this group and the overwhelming presence of low-SNR signals, these events were interpreted as small transverse crack events such as single fibre fractures, transverse matrix micro-cracking, and fibre-matrix interfacial slip. Frequency spectra taken from this region of the dataspace showed marked peaks at 95, 230, and 590 kHz (see Figure 5a).

There were, however, two exceptions in these regions, indicated by arrows in Figure 4:

(a) A 'small' event at $x = 144.6\,mm$ of specimen U7 with $z = 0.46$ shown in Figure 6a.
(b) A 'large' event at $x = 148.1\,mm$ of specimen U9 with $z = 0.03$ shown in Figure 6b, and preceded by three smaller events with similar $z \approx 0$ around $x \approx 151mm$.



(a) *Event (a),* $z = 0.46$, $x = 144.6\,mm$ **(b)** *Event (b),* $z = 0.03$, $x = 148.1\,mm$

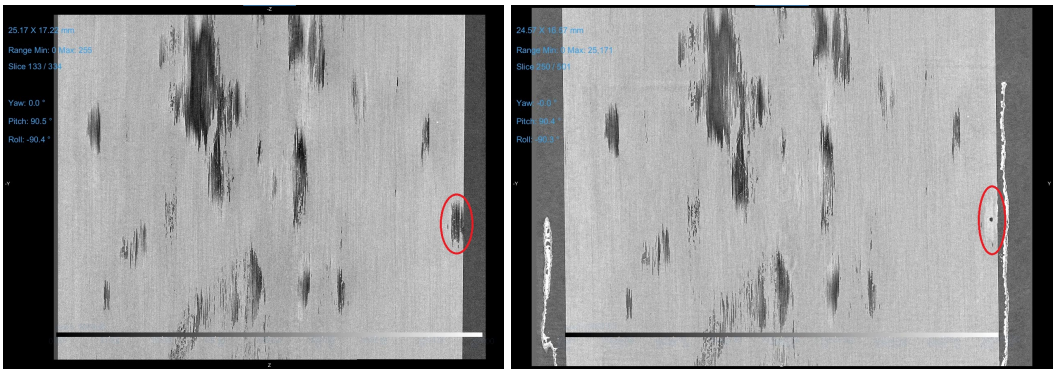
Fig. 6. *Waveforms recorded during events (a) and (b) in the scanned regions of the specimens.*

Literature suggests [16–18] that low-frequency AEs are attributed to matrix-related damage, and high frequencies come from fibre breakage. Therefore, event (a) was interpreted as an

interface failure between the layers of unidirectional prepreg because of its low-frequency content with peaks at around 35, 60, and 95 kHz. Event (b), on the other hand, was interpreted as a large fibre-bundle fracture event as the signal was 'loud' enough to saturate the data acquisition system, and the frequency spectrum (although somewhat unreliable because clipping in the time domain introduces spectral artifacts) showed a strong peak at 95 and contributions at 230 kHz.

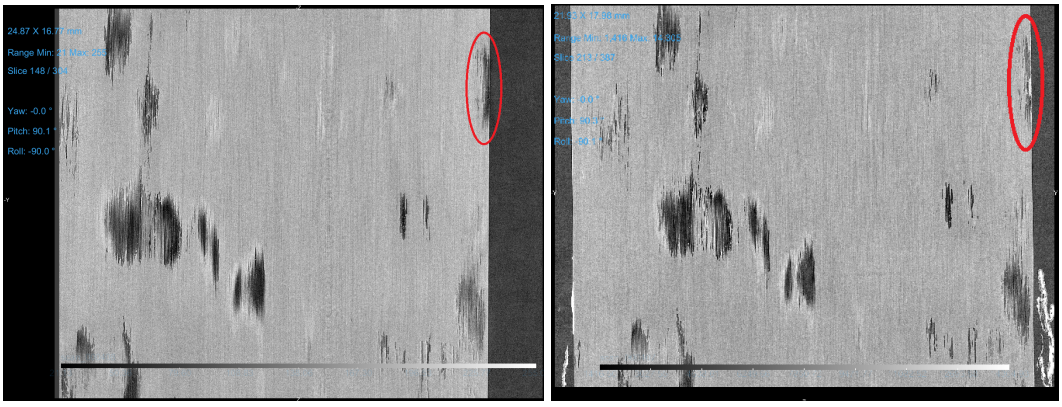
3.2 Validating micro-CT scans

The regions CT-scanned before and after tensile testing are marked by red rectangles in Figure 4, and the micro-CT scans are in good agreement with the interpretation of the AE monitoring data proposed in the previous section.
The scans performed at the mid-section of both specimens (centered at x = 125 mm) showed no visible changes before and after tensile testing, which agrees with the AE monitoring framework that detected only 'small cracks' ($z < 0$) in these regions - assuming they were below the visible resolution of the scans ($\sim 20\ \mu m$).
The scans of the top region are shown in Figures 7 and 8, and permanent changes are visible. The locations agree with the locations of events **(a)** and **(b)** respectively.



(a) *Before tensile testing* **(b)** *After tensile testing*

Fig. 7. *Specimen U7 micro-CT frames (IV, $x = 147mm$) before and after tensile testing.*



(a) *Before tensile testing* **(b)** *After tensile testing*

Fig. 8. *Specimen U9 micro-CT frames (IV, $x = 147mm$) before and after tensile testing.*

4. Discussion and Conclusions

The results demonstrate that unsupervised machine learning applied to normalized frequency spectra can effectively characterize AE waveforms and produce damage maps consistent with micro-CT observations.

This work is licensed under CCBY 4.0

Media and Publishing Partner NDTnet

Limitations, however, do exist. The first one is the openness to interpretation and current lack of verification on a large validating dataset. One simple way to overcome this is to repeat the experiment as-is on a larger batch of specimens to exclude possible coincidental correlation between AE monitoring results and micro-CT scans more confidently.
Another limitation is that including only the first six principal coordinates in the analysis may be excluding relevant spectral information that is hidden in higher-order components. On the other hand, including more higher-order principal components can increase the influence of noise and lead to a higher required complexity of the neural network without necessarily improving its final performance. In this work, it was observed that the interpretability of the damage maps remained high and consistent up until 8 principal components, and then decreased beyond that as low-amplitude AEs started to be differentiated by the algorithm, potentially on the basis of noise, in regions where no damage was visible in the CT scans. The number of principal components included in the analysis remains a tunable parameter in this method.
Finally, the sensors' frequency response function directly shapes the normalized spectra, and two sensors of the same model can still differ meaningfully in coupling. As a result, the latent variable z may reflect sensor-specific spectral characteristics in addition to source mechanism differences, and caution is advised when comparing z-values across experiments conducted with different sensors or coupling conditions.
The main strength of the proposed method lies in its potential for generalizability. Since it is based on unsupervised machine learning algorithms, it can be applied without modification to unknown AE waveform datasets and automatically infer data-space topology and differentiation of points within that space. No prior knowledge is needed about the fracture mechanics of observed damage, characteristic signal frequencies, or sensor transfer functions to automate AE differentiation. Instead, during training, the algorithm independently explores the data-space obtained using a certain type of sensor (which, in the case at hand, takes about 10 000 waveform signals and 1 minute using an average-speed computer), and AE character is inferred by the algorithm without prior knowledge of the data structure. An additional step of post-processing is subsequently required to interpret and contextualize the characterized signals, e.g., by using knowledge about strain and external events.
Results are expected to be directly comparable, with due caution, across experiments without retraining, provided that a few conditions are met, i.e. if the same sensors, type of coupling gel, sensor clamping force, and specimen boundary and loading conditions are used. The method's generalizability to comparable but different setups without retraining, on the other hand, will be addressed in future work.

Acknowledgments

This research has been conducted within the scope of 'MetaShip' project, which has been made possible by the financial contribution of COMMIT. The author furthermore gratefully acknowledges Ellen Meijvogel-de Koning for conducting the micro-CT scans of the composite coupon specimens.

References

[1] A. Huijer, C. Kassapoglou, and L. Pahlavan. "Acoustic emission monitoring of carbon fibre reinforced composites with embedded sensors for in-situ damage identification". In: *Sensors* 21.20 (2021), p. 6926.

[2] F. Riccioli et al. "Corrosion-fatigue damage identification in submerged mooring chain links using remote acoustic emission monitoring". In: *Marine Structures* 98 (2024), p. 103685.

[3] C. Muir et al. "Damage mechanism identification in composites via machine learning and acoustic emission". In: *npj Computational Materials* 7.1 (2021), p. 95.

This work is licensed under CCBY 4.0

Media and Publishing Partner

[4] N. Godin, P. Reynaud, and G. Fantozzi. "Challenges and limitations in the identification of acoustic emission signature of damage mechanisms in composites materials". In: *Applied sciences* 8.8 (2018), p. 1267.

[5] J. Rzeczkowski. "APPLICATION OF AN UNSUPERVISED DEEP LEARNING FRAMEWORK FOR ACOUSTIC EMISSION-BASED CHARACTERIZATION OF DELAMINATION PROCESS IN COMPOSITE LAMINATES". In: *Composites Science and Technology* (2026), p. 111507.

[6] ASTM International. *ASTM D3039 / D3039M-17: Standard Test Method for Tensile Properties of Polymer Matrix Composite Materials*. `https://www.astm.org/d3039_d3039m-17.html`. West Conshohocken, PA, USA. 2017.

[7] Toray Advanced Composites. *Toray 2510 Epoxy Prepreg System Technical Data Sheet*. `https://www.toraycma.com/wp-content/uploads/2510-Prepreg-System.pdf`. Accessed: 2025-04-01. 2021.

[8] R. Talreja. "Studies on the failure analysis of composite materials with manufacturing defects". In: *Mechanics of composite materials* 49.1 (2013), pp. 35–44.

[9] MISTRAS Group. *WSα Sensor*. `https://www.physicalacoustics.com/content/literature/sensors/Model_WSa.pdf`. Accessed: 2026-03-16. 2013.

[10] R. A. Johnson, D. W. Wichern, et al. *Applied multivariate statistical analysis*. Prentice hall Upper Saddle River, NJ, 2002.

[11] D. Baccar and D. Söffker. "Identification and classification of failure modes in laminated composites by using a multivariate statistical analysis of wavelet coefficients". In: *Mechanical Systems and Signal Processing* 96 (2017), pp. 77–87.

[12] M. Sakurada and T. Yairi. "Anomaly detection using autoencoders with nonlinear dimensionality reduction". In: *Proceedings of the MLSDA 2014 2nd workshop on machine learning for sensory data analysis*. 2014, pp. 4–11.

[13] I. Goodfellow, Y. Bengio, and A. Courville. *Deep Learning*. `http://www.deeplearningbook.org`. MIT Press, 2016.

[14] A. Huijer, C. Kassapoglou, and L. Pahlavan. "Acoustic emission source localization in fiber-reinforced composites based on multimodal dispersion compensation of guided waves". In: *STRUCTURAL HEALTH MONITORING 2023* (2023).

[15] J. L. Rose. *Ultrasonic guided waves in solid media*. Cambridge university press, 2014.

[16] P. J. De Groot, P. A. Wijnen, and R. B. Janssen. "Real-time frequency determination of acoustic emission for different fracture mechanisms in carbon/epoxy composites". In: *Composites science and technology* 55.4 (1995), pp. 405–412.

[17] C. Ramirez-Jimenez et al. "Identification of failure modes in glass/polypropylene composites by means of the primary frequency content of the acoustic emission event". In: *Composites Science and Technology* 64.12 (2004), pp. 1819–1827.

[18] R. Gutkin et al. "On acoustic emission for failure investigation in CFRP: Pattern recognition and peak frequency analyses". In: *Mechanical systems and signal processing* 25.4 (2011), pp. 1393–1407.

This work is licensed under CCBY 4.0

Media and Publishing Partner NDTnet