

A generic framework for application of machine learning in acoustic emission-based damage identification

Abhishek Kundu¹, Shirsendu Sikdar², Mark Eaton¹, and Rukshan Navaratne³

¹ Cardiff School of Engineering, Cardiff University, The Parade, Queen's Building, Cardiff, CF24 3AA, UK.

KunduA2@cardiff.ac.uk,

WWW home page:

<https://www.cardiff.ac.uk/people/view/364404-kundu-abhishek>

² Institute of Fluid-Flow Machinery, Polish Academy of Sciences, 14, Fiszerka Street, Gdansk 80-231, Poland.

³ University of South Wales, Treforest Campus, Pontypridd, CF37 1DL, UK.

Abstract. Advanced non-destructive monitoring scheme is necessary for modern-day lightweight composite structures used in aerospace industry, due to their susceptibility to barely visible damages from minor impact loads. Acoustic emission (AE) based monitoring of these structures has received significant attention in the past few years primarily due to their possibility of use in operating structures under service loads. However, localization and characterization of damages using AE is still an open area of research. The exploration of the space of signal features collected by a distributed sensor network and its reliable mapping to damage metrics (such as location, nature, intensity) is still far from conclusive. This problem becomes more critical for composite structures with complex features/geometry where the localized effects of discontinuity in geometric or mechanical properties do not make it appropriate to rely on simple signal features (such as time difference of arrival, peak amplitude, etc.) to identify damage. In this work, the AE signal features (which are spatially and temporally correlated) have been mapped to the damage properties empirically with a training dataset using meta-modeling techniques. This is used in the online monitoring phase to infer the probabilistic description of the acoustic emission source within a hierarchical Bayesian inference framework. The methodology is tested on a carbon fibre composite panel with stiffeners that is subjected to impact and dynamic fatigue loading. The study presents a generalized machine learning-based automated AE damage detection methodology which both localizes and characterizes damage under varying operational loads.

Keywords: acoustic emission, feature extraction, machine learning, damage characterization, gaussian process

1 Introduction

Acoustic Emission (AE) is the result of changes within a structure or material which results from mechanical phenomenon of plastic deformation, friction and rubbing or crack growth which lead to elastic wave propagation in a structure. The propagating AE waves generate minute surface-displacements when they travel along a surface. These displacements can then be converted to the voltage response by using the piezoelectric AE-sensors. These transient signals can be analysed in real-time using most modern AE systems and can also be stored and digitised to allow further signal processing. The triangulation methods can be effectively used to predict the damage source-locations in the target structure by using a distributed network of AE-sensors that detects the released energy from an AE-source, as in the case of determining the epicentre of an earthquake in seismology. A schematic representation of the AE based SHM process is described in Fig. 1 which clearly indicates that any changes in the geometry or material will affect the resultant recorded signal.

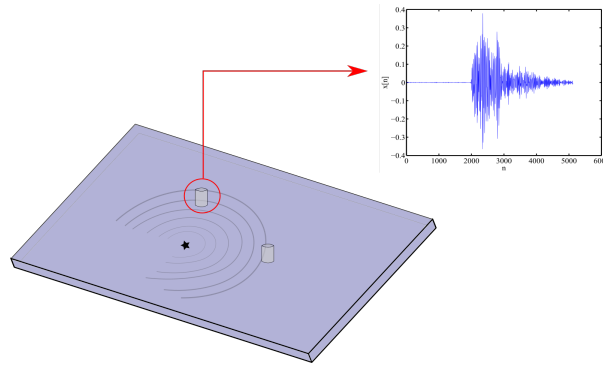


Fig. 1. AE measurement process and typical AE signal recorded at a sensor.

Accurate localization of damage-source location using AE signals in complex structures is significantly challenging as the proper assessment of boundary level uncertainties in the mathematical model is complex in nature and associated with substantial computational overhead [1]. The loss in accuracy in complex structures is mainly occurred due to the failure in accurate determination of time-of-arrival of the AE-signal and due to the overly involved in the simplified representation of the wave propagation path and velocity during source-location calculations [2,3]. The present study aims to overcome the above limitations using a novel approach which calibrates and trains the damage-model using a compound-correlation-metric among the AE-signals registered from the predefined network of AE-sensors. This is then utilized to create a predictive model which performs inverse damage characterization using the Bayesian inference framework [4,5].

This study aims to overcome the above limitations using a novel approach which trains and calibrates the damage model using a compound correlation metric between the signals recorded with the distributed sensory network. When only using the arrival time of the signal at the individual sensors (and the difference in arrival time) as the basis of training the model, a large portion of the signal data is not being utilized fully. Additionally error is incurred in the ad hoc definitions of signal threshold values which are used to calculate arrival time. Moreover, it is strongly based on the assumption of the existence of a travel path between the damage and the sensor location (this can be considered as an underlying regularization) which might often not be the case. The proposed methodology, in contrast, constructs a compressed representation of the full signal characteristics using a projected correlation metric (as discussed in section 3) which utilizes the full signal characteristics to infer the source of the incoming waves. The Gaussian process based surrogate regression approach explicitly accounts for the uncertainty in lack of training data and/or the error incurred using the process described in section 4. This is followed by section 5 which gives the description of the test rig on which the AE source localization algorithm is applied, section 6 where the main findings of the study has been presented and the conclusion and future works is included in section 7.

2 Acoustic emission source location

Time of arrival (TOA) based direct methods for determination of arrival time (and hence the difference in arrival time) of the signal at different sensors are widely used in all commercial AE systems. In this method, a predetermined signal threshold level is defined and the point at which the transducer voltage response exceeds this level is taken to be the signals arrival time. It is, however, possible for signal to be present prior to the first threshold crossing and attenuation of signal amplitude can mean variation in signal arrival time determination. A wide range of approaches have been developed in an attempt to improve the arrival time determination compared with the traditional threshold crossing technique. A range of frequency based techniques including filtering [6], cross-correlation [7] and wavelet transforms [8,9] have been investigated, however, statistical approaches based on 6th order statistical moments [10] and the Akaike Information Criteria (AIC) [11,12] have been shown to be more reliable. The AIC approach in particular has been demonstrated to be very robust across a range of materials and structures [13].

Standard AE location algorithms assume a single, constant, wave propagation speed, however, in composite materials the wave speed is seen to vary with propagation direction and is dependent on fibre orientations within the layup used. Several researchers have tried to address this challenge and some success has been by extending the traditional time of arrival optimization scheme to include a variable wave speed dependent on propagation direction [14,15]. Ciampa and Meo [16] adopted a novel approach whereby closely spaced sensor pairs were used to reduce the number of unknown propagation velocities in a set of

simultaneous non-linear equations that describe the source position. An iterative Newton approach was adopted to solve the unknowns in the equations and therefore yield an estimate of the source position without prior knowledge of the wave speeds in the material. Despite these advances achieved in anisotropic monolithic materials, none of the approaches are capable of accounting for geometric complexities such as access holes, curvatures and thickness changes that commonly occur in industrial structures. Alternatively a mapping approach has been proposed and rigorously validated [17,18,13,19,20] in which a structure is mapped using artificial AE sources (such as a H-N source [21,22]) to derive an empirical relationship between the known source position and the resultant arrival times at an array of sensors. This relationship can then be used to determine the source origin for a set of measured arrival times. The empirical nature of the approach inherently accounts for all material and structural complexity such as anisotropy and geometric features.

The previously discussed approaches are deterministic and do not consider the uncertainty in the measurements and calculations performed. This is particularly relevant to their application to industrial environments where uncertainty is seen to increase, with varying operating conditions such as temperature affecting wave propagation and therefore reducing reliability. To account for the uncertainty that can be experienced in an industrial environment researchers have begun to adopt probabilistic approaches. Schumacher *et al* [23] developed an approach based on Bayesian statistics for AE source location in a reinforced concrete beams that accounts for uncertainties and errors that exist within the measurement and calculation process. A simplified model for the concrete beam was developed, in which the mean of the wave slowness, the standard deviation of the wave slowness, the event time and the standard deviation of the observed arrival times are represented as prior probability density functions (PDF). The initial PDF of each parameter was then refined using experimental data collected from H-N sources at known positions on the beam surface. The refined model could then be used to predict the most likely position of any subsequent AE sources. The approach reduced the mean error of 22 arbitrarily located H-N sources from $\sim 40\text{mm}$ down to $\sim 30\text{mm}$. Further work by Zarate *et al* [24] developed a Bayesian framework based on a ray tracing model of AE wave propagation in liquid filled storage tanks. The approach allowed structure borne and water borne wave paths to be considered. Using a Markov Chain Monte Carlo (MCMC) method to sample the posterior distribution of the source position in x and y coordinates the most probable source position could be determined. Both of these probabilistic approaches are limited to homogeneous materials and simple geometries, i.e. direct and uninterrupted wave paths. The study of signal characteristics mapped to the damage location has been undertaken by the present authors [25,26], whereby Bayesian inference has been used for source localization in the prediction stage. This paper is intended to highlight the main aspects of that proposed approach and extend it further to the problem of damage characterization.

This main aim to the present work is to combine the Bayesian probabilistic mapping of the source of AE to the correlation characteristics of the signal collected at the distributed sensor network. This would not only capture the essential information pertaining to the phase difference and attenuation of elastic waves travelling different distances over the surface of the composite structures to reach individual sensors but also include the complex effects of boundary reflection within the signal correlation characteristics. The uncertainty due to the measurement noise and experimental errors would be explicitly accounted for in the probabilistic model and conditioned on the training data generated with H-N sources.

3 Important signal characteristics for damage identification

3.1 Compressed cross-correlation signal features for AE

The signal features recorded at different locations on the structure being investigated can be mapped to the damage characteristics to obtain a data-driven identification framework. The main challenge with this approach is that the sensors data can be affected by various ambient and operational factors, such as changes in dispersion behavior, sensor-structure coupling characteristics, amongst others. Hence a trained data-driven model which does not account for the changes in baseline environmental, operational conditions would provide erroneous damage identification with changes in the above conditions.

One way to compensate for this is to focus on sensor data in relation to each other. Specifically, the correlation measure between the signals observed at different sensor locations in a distributed sensory network is quite useful because the normalized correlation measures are invariant to measurement noise and can compensate for changes in ambient condition, such as temperature sensitivity.

If we consider n_s sensors distributed on the structure under study, the collected signal is represented as

$$\mathbf{X} = \{\mathbf{x}_i : \mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{in})^T \quad \forall i = 1, \dots, n_s\} \quad 1$$

where each sample x_{ij} has been collected at sensor i at n discrete time points $j = 1, \dots, n$. An initiation of crack or a growth in crack size is accompanied by a packet of ultrasonic wave which propagates radially outwards from its point of origin along the plane of the composite structure. Each sensor captures signal \mathbf{x}_i as show in Fig. 1. Additionally reflection from boundaries and/or other geometrical discontinuities (such as holes) result in a complicated propagation characteristics. It is expected that the accumulated data from the distributed sensory network would contain distinguishing features (statistics) which would map the detected waveforms uniquely to the source location.

The correlation matrix $\bar{\mathbf{R}}$ between the signals collected at each of the n_s sensors is defined as

$$\bar{\mathbf{R}}_\tau = \begin{bmatrix} \mathbf{R}_{11}(\tau) & \mathbf{R}_{12}(\tau) & \cdots & \mathbf{R}_{1n_s}(\tau) \\ \mathbf{R}_{21}(\tau) & \mathbf{R}_{22}(\tau) & & \mathbf{R}_{2n_s}(\tau) \\ \vdots & & \ddots & \vdots \\ \mathbf{R}_{n_s 1}(\tau) & \mathbf{R}_{n_s 2}(\tau) & \cdots & \mathbf{R}_{n_s n_s}(\tau) \end{bmatrix} \quad 2$$

where each element $\mathbf{R}_{ij}(\tau)$ of the correlation matrix is defined in the continuous time domain as

$$R_{ij}(\tau) = \int_{-\infty}^{\infty} x_i(t)x_j(t+\tau)dt \quad 3$$

for real-valued signal components. For L^1 integrable signals, i.e. $\int_{-\infty}^{\infty} |x(t)| dt < \infty$ the Fourier transform of the signal exists and is related to the terms in $\bar{\mathbf{R}}$ matrix. In Eqn. 3, when $i = j$, the terms \mathbf{R}_{ii} , i.e. the diagonal terms in the matrix $\bar{\mathbf{R}}_\tau$, give the autocorrelation measure of the signals at each sensor.

For discrete time signals the cross-correlation for real-valued signals is defined as

$$\mathbf{R}_{ij}(\tau) = \mathbb{E} [(\mathbf{x}_i[n] - \boldsymbol{\mu}_i)(\mathbf{x}_j[n + \tau] - \boldsymbol{\mu}_j)] \quad \forall i, j = 1, \dots, n_s \quad 4$$

The cross-correlation is often normalized by the respective auto-correlation functions at zero lag such that

$$\mathbf{r}_{ij}(\tau) = \frac{\mathbf{R}_{ij}(\tau)}{\sqrt{\mathbf{R}_{ii}(0)\mathbf{R}_{jj}(0)}} \quad \text{where } |\mathbf{r}_{ij}(\tau)| \leq 1 \quad \forall i, j, \tau \quad 5$$

and we denote the normalized version of $\bar{\mathbf{R}}_\tau$ as $\bar{\mathbf{r}}_\tau$ where each element of the matrix in Eqn. 2 is normalized as per Eqn. 5. It is important to note that $\mathbf{R}_{ij}(\tau) = \mathbf{R}_{ji}(-\tau)$ and $\mathbf{R}_{ij}(\tau) = \mathbf{R}_{ij}(-\tau)$, hence \mathbf{r}_τ is a symmetric matrix for all τ . The normalized correlation coefficient is utilized here to balance the varying intensity of the AE source. The mapping which links training dataset to the source location assumes that the source intensity is normalized across the entire training dataset which is achieved with the normalized correlation coefficient.

The correlation matrix captures the essential information regarding the correlation of the signals captured using the distributed sensor network which contains essential information regarding the delay or phase of the arriving signals. The matrix $\bar{\mathbf{R}}_\tau$ is constructed over the interval $-\tau_s \leq \tau \leq \tau_s$ such that each of its elements $\mathbf{R}_{ij}(\tau)$ or the normalized cross-correlation $\mathbf{r}_{ij}(\tau)$ is a vector of dimension $2\tau_s + 1$. Since the correlation matrix is symmetric, only the upper triangular part is considered in a matrix representation

$$\bar{\mathbf{r}}_{\Delta_k, \tau_s} = [\mathbf{r}_{\tau_s, i} : \mathbf{r}_{\tau_s, i} \in \mathbb{R}^{2\tau_s+1} \text{ and } -\tau_s \leq \tau \leq \tau_s \quad \forall i \in \mathcal{I}] \quad 6$$

where \mathcal{I} is the set of indices associated with the ordering of the elements of the upper triangle of \mathbf{r}_τ at some instance τ and $\mathbf{r}_{\tau_s, i}$ is a vector which contains the cross-correlation measure over the interval $\pm\tau_s$. It is important to note that the subscript Δ_k of the matrix $\bar{\mathbf{r}}_{\Delta_k, \tau_s}$ $k = 1, \dots, n_t$ denotes normalized correlation

data evaluated for the k -th test point out of a total of n_t training sets. Thus the full set of training correlation data is given as

$$\bar{\mathbf{r}}_{\Delta, \tau_s} = [\bar{\mathbf{r}}_{\Delta_1, \tau_s}, \bar{\mathbf{r}}_{\Delta_2, \tau_s}, \dots, \bar{\mathbf{r}}_{\Delta_{n_t}, \tau_s}] \quad \text{where} \quad \bar{\mathbf{r}}_{\Delta_k, \tau_s} \in \mathbb{R}^{2\tau_s+1 \times n_{\mathcal{I}}}, \quad k = 1, \dots, n_t \quad 7$$

where $n_{\mathcal{I}}$ is the cardinality of the set \mathcal{I} .

The objective here is to obtain a map between the normalized correlation matrix representing the signal data for each of the n_t training sets and the corresponding coordinates of the acoustic source. However, the computational demands of modelling the large predictor matrix $\bar{\mathbf{r}}_{\Delta, \tau_s}$ is substantial and would make the subsequent real-time model prediction infeasible. Thus we seek an optimal reduced basis which can be used to represent the data for each training set such that

$$\begin{aligned} \text{find } \mathcal{T} : \mathbb{R}^{2\tau_s+1 \times n_{\mathcal{I}}n_t} &\rightarrow \mathbb{R}^{n_r \times n_{\mathcal{I}}n_t} \quad \text{where} \quad \mathcal{T}(\bar{\mathbf{r}}_{\Delta_i, \tau_s}) = \boldsymbol{\alpha}_{i, \tau_s} \quad \forall i = 1, \dots, n_{\mathcal{I}}n_t \\ \text{such that } \Phi_{\Delta, \tau_s} &= \arg \inf_{\Phi_{\Delta_j, \tau_s}, \forall j} \|\Phi_{\Delta, \tau_s}[\boldsymbol{\alpha}_{1, \tau_s}, \dots, \boldsymbol{\alpha}_{n_{\mathcal{I}}n_t, \tau_s}] - \bar{\mathbf{r}}_{\Delta_i, \tau_s}\|_F \end{aligned} \quad 8$$

where $n_r < 2\tau_s + 1$ and $\boldsymbol{\alpha}_{i, \tau_s} \in \mathbb{R}^{n_r}$. Thus the transformation \mathcal{T} expresses the normalized correlation terms using an optimal basis on which the solution would be projected and an additional orthonormality constraint of $\Phi_{\Delta, \tau_s}^T \Phi_{\Delta, \tau_s} = \mathbf{I}$ can be imposed on the basis. This optimal basis can be informed by the left singular eigenvectors of the training correlation data contained in $\bar{\mathbf{r}}_{\Delta, \tau_s}$ or the principal eigenvectors associated with $\bar{\mathbf{r}}_{\Delta, \tau_s}^T \bar{\mathbf{r}}_{\Delta, \tau_s}$. The compressed correlation data is expressed in terms of this optimal basis as

$$\underline{\boldsymbol{\alpha}}_{\tau_s} = \{\boldsymbol{\alpha}_{1, \tau_s}, \dots, \boldsymbol{\alpha}_{n_t, \tau_s}\} = \Phi_{\Delta, \tau_s} \quad \text{where} \quad \bar{\mathbf{r}}_{\Delta, \tau_s}^T \quad \text{where} \quad \underline{\boldsymbol{\alpha}}_{\tau_s} \in \mathbb{R}^{n_r \times n_{\mathcal{I}}n_t} \quad 9$$

The matrix $\underline{\boldsymbol{\alpha}}_{\tau_s}$ consists of a collection of $n_{\mathcal{I}}n_t$ vectors each of dimension n_r , thus for each of the n_t training points we get a set of $n_{\mathcal{I}}$ vectors which contains the compressed information.

Next we introduce the linear map $vec(\cdot) : \mathbb{R}^{n \times k} \rightarrow \mathbb{R}^{nk}$ which transforms a collection of k vectors each of dimension n into a single vectors of dimension nk (where each of the k vectors are stacked end to end). The matrix $\bar{\mathbf{r}}_{\Delta, \tau_s}$ is transformed as

$$\underline{\boldsymbol{\alpha}}_v = \{vec(\boldsymbol{\alpha}_{1, \tau_s}), \dots, vec(\boldsymbol{\alpha}_{n_t, \tau_s})\} \quad \text{where} \quad \underline{\boldsymbol{\alpha}}_v \in \mathbb{R}^{n_r n_{\mathcal{I}} \times n_t} \quad 10$$

such that each of the $n_r n_{\mathcal{I}}$ vector in $\underline{\boldsymbol{\alpha}}_v$ corresponds to the compressed data collected at the distributed sensor network on the structural system.

The problem then defined defined in terms of finding the map $\mathcal{G} : \boldsymbol{\alpha}_{v, xy} \rightarrow r_{xy}$ where r_{xy} is a vector of the Euclidean distance of AE source from each sensor in the distributed sensory network and the subscript (xy) denotes the Cartesian coordinates of the AE source. The problem is posed as the following multivariate least square regression problem

$$\boldsymbol{\gamma} = \arg \inf_{\boldsymbol{\gamma} \in \mathbb{R}^{n_r n_{\mathcal{I}}}} \|\boldsymbol{\alpha}_{v, xy} \boldsymbol{\gamma} - r_{xy}\|_2 \quad 11$$

However, for each training run, there is an uncertainty around the measured signal. We construct a scalar error indicator following from Eqn. 11 as

$$\varepsilon_r = \|\hat{r}_{xy} - r_{xy}\| \quad \text{where} \quad \hat{r}_{xy} = \alpha_{v,xy} \gamma \quad 12$$

where \hat{r}_{xy} is the approximate identified location of the acoustic source associated with the training data where the known true source is at r_{xy} distance from the sensors. Eqn. 12 gives the error associated with the regression map $\mathcal{G} : \alpha_{v,xy} \rightarrow r_{xy}$. Thus the error indicator ε_r for the training dataset is mapped on the spatial domain using a Gaussian process surrogate, discussed in the following section, based on which the probabilistic prediction of the source location is performed based on the observed signal data at the distributed sensory network.

3.2 Discrete scale-time decomposition of cross-correlation

The cross-correlation based signal compression can be further filtered to localize specific frequency components in the time domain. This gives the advantage of accurately identifying the arrival of specific propagation modes at different sensor locations. Hence a better understanding of the dispersion behavior is obtained by this. Additionally, the proportion of energy concentrated in different wave modes can be calculated from the scale-time decomposition of the signals.

Using discrete and continuous wavelet transform for signal characterization and feature extraction is an established field of study. We briefly discuss the fundamental aspects of the principle of wavelet transform here and its application to the cross-correlation metric. Assuming a mother wavelet of the form $\Psi_{s,\Delta}(t) = \frac{1}{\sqrt{s}} \Psi\left(\frac{t-\Delta}{s}\right)$ and applied to a signal $x(t)$ we get

$$X_i^{s,\Delta} = \int_t x_i(t) \Psi_{s,\Delta}^*(t) dt \quad 13$$

where s and Δ are the scale and time parameters respectively. It is to be noted that the dual of this relationship exists, which enables the reconstruction of the original signal from the inverse wavelet transform. Applying the wavelet transform to the normalized cross-correlation measure (defined in Eqn. 5) results in

$$\mathbf{r}_{ij}^{s,\Delta} = \int_\tau \mathbf{r}_{ij}(\tau) \Psi_{s,\Delta}^*(\tau) d\tau = \frac{1}{\sqrt{\mathbf{R}_{ii} \mathbf{R}_{jj}}} \int_t \int_\tau x_i(t) x_j(t + \tau) \Psi_{s,\Delta}^*(\tau) d\tau dt \quad 14$$

Simplifying the expression in the Eqn. 14 and applying change of variables we obtain

$$\mathbf{r}_{ij}^{s,\Delta} = \frac{1}{\sqrt{\mathbf{R}_{ii} \mathbf{R}_{jj}}} \int_t x_i(t - \Delta) X_j^{s,t} dt \quad 15$$

This facilitates rapid calculation of the $\mathbf{r}_{ij}^{s,\Delta}$ as features using which the localization and characterization of AE sources can be made.

The proportion of energy concentrated in the propagating wave modes of the AE signals recorded at a particular sensor location (say i) is given in a scale-time sub-domain $s \in s_h$; $\Delta \in \Delta_h$ as

$$E_i^{s_h, \Delta_h} = \int_{s_h} \int_{\Delta_h} \left| X_i^{s, \Delta} \right|^2 ds d\Delta \quad 16$$

where s_h and Δ_h are sub-domains defined the scale and time axis respectively. The energy concentration in a scale-time sub-domain $[s_h, \Delta_h]$ of the wavelet transformed correlation measures is given by

$$E_{ij}^{s_h, \Delta_h} = \int_{s_h} \int_{\Delta_h} \left| \mathbf{r}_{ij}^{s, \Delta} \right|^2 ds d\Delta \quad 17$$

Here $E_{ij}^{s, \Delta}$ gives a quantitative estimate of the product of energy concentration in the correlated signals. For the case when $i = j$ and $\Delta = 0$, Eqn. 17 gives the energy concentration in the scale-time domain of the signal at sensor i . The quantity $E_{ij}^{s, \Delta}$ would be incorporated as features in the Gaussian process surrogate to facilitate the data-driven framework for AE source characterization in the

4 Gaussian process regression with signal features

In the previous section, the error indicator ε_r is mapped to the (x, y) coordinates of the AE source as $f_\varepsilon: [0, l_x] \times [0, l_y] \ni (x, y) \mapsto \varepsilon_r \in \mathbb{R}^+$. In this section, we introduce a Gaussian process model [27] to compute the map f_ε .

The coordinates (x, y) of the AE source over the domain $[l_x, l_y]$ serves as the parameter space or the input space over which the error indicator ε_r is trained. Here it is assumed that the observed error output vector $[\varepsilon_{r_1}(x, y), \dots, \varepsilon_{r_{n_t}}(x, y)]$ are realisations of a Gaussian stochastic process with parametrized mean and covariance functions. The model structure is thus expressed as

$$\varepsilon_r(x, y) = \mathbf{h}(x, y)^\top \boldsymbol{\beta} + Z(x, y) \quad 18$$

where $\mathbf{h}(x, y)$ is a vector of known functions and $\boldsymbol{\beta}$ is an unknown hyperparameter to be estimated from the data. The choice of $\mathbf{h}(\boldsymbol{\xi})$ is an active research area [28, 29]. The function $Z(\cdot)$ is a stochastic process with mean zero and covariance function

$$\text{Cov}(Z((x, y)), Z(x', y')) = \sigma_z^2 C((x, y), (x', y')) \quad 19$$

where $C(\cdot, \cdot)$ is a correlation function and σ_z^2 is the process variance, a hyperparameter that can also be estimated from the data. In order to choose a valid positive-definite correlation function, we consider products of one-dimensional correlations[30] with functional forms

$$C((x, y), (x', y')) = \exp \left\{ -b_i |x - x'|^2 - b_y |y - y'|^2 \right\} \quad 20$$

where $b_i > 0$ for all i with $x, x' \in [0, l_x]$ and $y, y' \in [0, l_y]$. The above correlation function is infinitely differentiable which is convenient when Gaussian processes are used to model not only the code output, but also its derivatives [31]. The vector of smoothness hyperparameters $\mathbf{b} = (b_1, b_2)^\top$ quantifies the rate at which the output varies over the spatial domain.

Let $\mathcal{D} = \{(x, y)_i, \varepsilon_r(x, y)_i\} | i = 1, \dots, n_t\}$ be a set of n_t training runs. Given this observed dataset, Bayes' theorem is used to estimate the hyperparameters as $\mathbb{P}(\beta, \sigma_{\mathbf{z}}^2, \mathbf{b} | \mathcal{D}) \propto \mathbb{P}(\mathcal{D} | \beta, \sigma_{\mathbf{z}}^2, \mathbf{b}) \mathbb{P}(\beta, \sigma_{\mathbf{z}}^2, \mathbf{b})$, where $\mathbb{P}(\beta, \sigma_{\mathbf{z}}^2, \mathbf{b} | \mathcal{D})$ is the posterior probability of the hyperparameters, $\mathbb{P}(\mathcal{D} | \beta, \sigma_{\mathbf{z}}^2, \mathbf{b})$ is the likelihood, $\mathbb{P}(\beta, \sigma_{\mathbf{z}}^2, \mathbf{b})$ is the prior of the hyperparameters, and $\mathbb{P}(\mathcal{D})$ is the marginal likelihood. A detailed derivation of prior-to-posterior analysis along with the hyperparameter estimation is given in [32,33].

The assumed Gaussian process prior on the code's output implies that the posterior distribution is also a Gaussian process. Once the hyperparameters are estimated, the mean of the posterior distribution approximates the output of ε_r^* at any point (x^*, y^*) on the physical domain. The variance of the posterior distribution quantifies the uncertainty that arises from having only a limited number of observations [34]. It can be shown that the posterior distribution is a Gaussian with the posterior mean and covariance function given in [27]. The detailed analysis is not repeated here for the sake of brevity.

The error indicator can hence be evaluated at any (x, y) (as predicted by the multivariate least square regression problem in Eqn. 11) and $\varepsilon(x, y)$ gives the associated uncertainty in the acoustic source location. The advantage offered by the above approach is that following the trained Gaussian process surrogate, the error around any predicted value can be computed directly from the surrogate in a computationally efficient manner. This enables real-time damage localization in in-situ operating structures.

The error $\varepsilon(x, y)$ constructed here is used in conjunction with the least square estimator (LSE) using the correlation measures (presented in section 3) gives a probabilistic prediction of the AE source location conditional on the training data. At the prediction stage the reduced correlation matrix, constructed using the optimal basis Φ_{Δ, τ_s} , is mapped to the AE source and the associated error is evaluated from the trained Gaussian process surrogate. Thus a robust probabilistic estimate of the AE source location is obtained using the correlation characteristics of the signal collected from the distributed sensor network.

5 Experimental setup

In order to develop and validate the above approach for damage localization, data was collected from a stiffened carbon fibre composite panel, representative of an aerospace structure.

5.1 Sample details

The manufactured stiffened panel is presented in Figure Fig. 2. The stiffeners were purchased from Easy Composites Ltd. (Staffordshire, UK) they consist of

a 90° L-shaped cross-section with laminate thickness of 3mm and cross-section dimensions of 25×25 mm and a length of 600mm. The stiffeners are made from

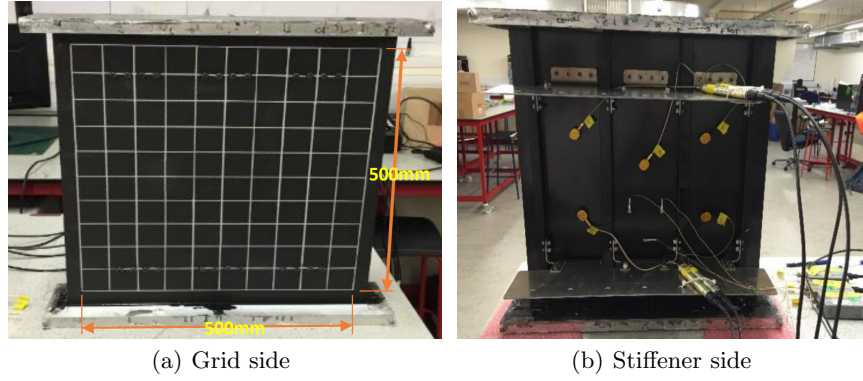


Fig. 2. Manufactured panel a) from stiffener side showing sensor positions and b) from skin side showing grid used for data collection.

2×2 twill high strength carbon fibres in an epoxy matrix with fibres aligned in the 0° and 90° directions. The skin was manufactured from 12 plies of Cytec MTM28/T800HB/200/42% 2×2 twill weave carbon fibre composite material with a $(0)_{12}$ layup and was cured in an autoclave in line with the manufacturers recommended cure cycle. Following curing the thickness was 2.85mm and the skin panel was cut to 550×600 mm using a water cooled diamond tipped cutting wheel. The skin and stiffeners were lightly abraded and degreased in preparation for bonding using permabond ET5429 adhesive. The final overall panel dimensions were 550×600 mm with the stiffeners running vertically. Two aluminium dummy ribs (representative of attachment to a wing rib for a composite wing skin) were attached at $1/4$ and $3/4$ height of the panel by drilling and bolting using 12 M4 bolts for each, as seen in 2(b). The top and bottom edges of the panel were potted into 20mm deep aluminium frames using Airtech TMR2001 high temperature laminating resin to allow application of a compressive load (not considered in this paper). A 500×500 mm grid with 50mm resolution was applied to the skin side of the panel (2(a)) and was used to aid the collection of training data.

5.2 Panel instrumentation

The panel was instrumented with five AE sensors as seen in 2(b), five McWade NS-3303 (300 kHz), the larger gold coloured sensors, and three Mistral Group Ltd. Nano30 sensors (300kHz), the smaller silver sensors. Four McWade sensors are arranged in a 275×175 mm square with the fifth placed centrally. The Nano30 sensors are arranged in a 75mm spaced triangular array. The sensor outputs are amplified by 40dB using a McWade PA3303 pre-amplifier for the

McWade sensors and a Mistral Group Ltd. 2/4/6 (20-1200kHz) pre-amplifier for the Nano30 sensors. A silicon adhesive (Loctite 595) was used to attach the sensors and provide a suitable acoustic couplant, the adhesive was allowed to cure for 24 hrs before any data acquisition or testing was undertaken and the correct coupling of the sensors was assessed using a H-N source [21,22]. The AE data was recorded with a Mistral Group Ltd. PCI-2 acquisition system using a 45dB threshold level. The detected signals were sampled at 5MHz for a duration of 1000 μ s: 600 μ s after the threshold crossing point and with a pre-trigger of 400 μ s. The time when the threshold crossing occurs for each test point has been recorded and shown in Fig. 3 and has been discussed in section 6.

5.3 Data acquisition

AE data was collected from the manufactured composite panel using a H-N source [21,22] to excite artificial AE waves. The H-N source is recognised as a standard reference source (ASTM E976) for AE testing and requires the fracture of a 0.5mm diameter 2H pencil lead against the sample surface at an angle of 30°. This is facilitated using a propelling pencil fitted with a plastic rocker the rocker is placed on the sample and the pencil rotated until the lead contacts the surface and then fractures. This results in minute elastic deformation of the surface under the tip of the pencil lead and when the lead fractures the elastic energy stored in the surface is rapidly released and excites a broadband elastic stress wave. This is highly representative of the rapid release of elastic energy that occurs when cracks and fractures grow in materials and hence is why it has been adopted as a standardised reference source for AE testing. The H-N source commonly excites a larger amplitude signal than a real fracture, however, the source mechanism is still representative and it has been shown to be a suitable artificial source for training a system for the detection and analysis of AE signals from real fracture events as seen in the Delta T Mapping techniques discussed above.

The data used in this work were collected using H-N sources performed at the nodes of the grid applied to the front of the stiffened panel. Ten H-N sources were conducted at each of the grid nodes within the 500 × 500mm grid shown in 2(a) and for each H-N source five AE signals were recorded and stored (one from each sensor).

6 Results and discussion

A sample of the collected sensor data is shown in Fig. 3 where each red dot signify the time (x-axis) of arrival of the wave at a sensor (5 sensors denoted by “Ch 1” to “Ch 5”) and its maximum amplitude (y-axis). A *hit* is recorded when the signal received at a particular sensor exceeds a preset threshold value. Fig. 3 shows the hits recorded simultaneously at all the five sensor locations. All the hits between the vertical blue lines on each channel indicate a test performed with the H-N source at a grid point (Fig. 2(a)) on the panel. Around 10 tests

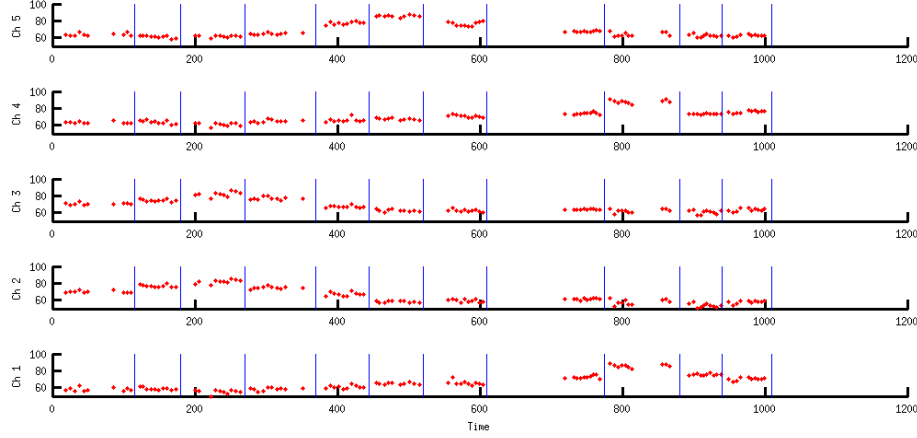


Fig. 3. Normalized correlation coefficients of the signal collected at the distributed sensor network (only the first 4 out of the total of 5 sensor signals shown) over discrete time steps $-100 \leq \tau_s \leq 100$

have been performed at each grid point due to which there are approximately 10 hits (red dots) between two consecutive blue lines. The red dots are aligned in time across all channels except for the small time difference associated with the waves travelling different distances from the source to reach each sensor. All the tests represented in Fig. 3 have been performed along the 11 test grid points at $x = 0$ (the bottommost line in the grid). This figure essentially represents only the time of arrival for the signal at the individual sensors, but the correlation measure between the sensors is shown in the next figure.

The optimal basis on which the correlation matrix is projected to obtain a compressed representation of the information collected at the sensors. A total of 30 basis functions have been used to approximate the terms of the correlation matrix. The basis functions are orthonormalized and have been calculated using the left eigenvectors of the correlation matrix using the singular value decomposition. The rapidly decaying singular value spectrum of the ensures that a good approximation is obtained with the reduced basis. The correlation matrix projected on these 30 optimal bases is termed as the reduced correlation matrix.

The accuracy of the identified AE source using the LSE is shown in Fig. 5. The black dots indicate the training points on the 10×10 grid of the composite panel while the red dots indicate the approximate identified acoustic source locations obtained using the least square mapping of the reduced correlation matrix to the vector of the distance of the acoustic source from the sensor network (as discussed in Eqn. 11). The error $\varepsilon_r(x, y)$ is given by the distance of the identified source locations (red dots) to the known position of AE (black squares).

This error indicator $\varepsilon_r(x, y)$ varies as per the accuracy of the fit and a Gaussian process surrogate is used to build a probabilistic error surface over the spatial domain as discussed in section 4. Fig. 5 gives the mean and standard de-

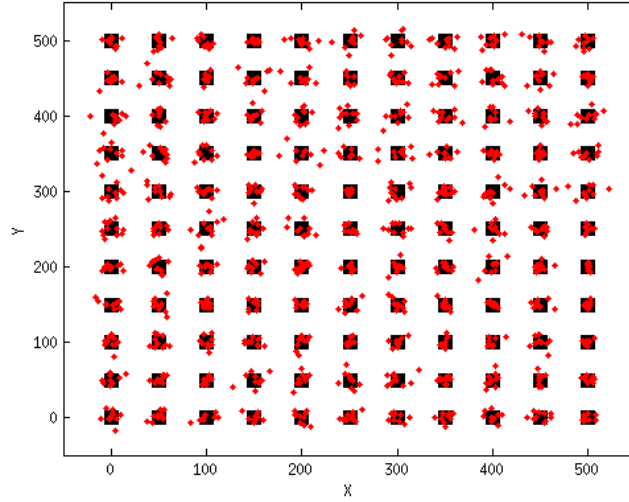


Fig. 4. The accuracy of fit of the training dataset based on the observed correlation matrix over the physical domain of the composite panel. The square blocks shown in black constitute the actual test grid while the red dots are LSE of the AE source.

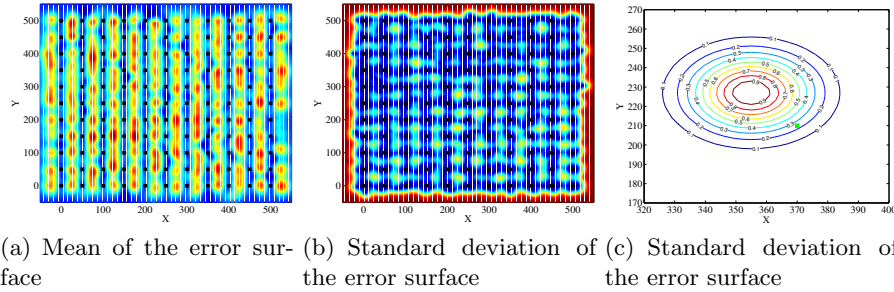


Fig. 5. The mean and standard deviation of the error response surface over the domain of the plate. The colormap of (a) varies between 0 – 45mm while for the standard deviation is between 0 – 20mm

variation of the posterior error surface conditional on the observations. This gives a distribution of the accuracy of the LSE as a function of the spatial coordinates. It can be verified that the mean error is almost zero at those training locations where the accuracy of the least square method is maximum. The locations where there are no training points show a high value of mean error and correspondingly high standard deviation of error which is an expected behavior.

Fig. 5(c) shows the probability distribution associated with an identified test point. The actual location of the acoustic source is shown as the green dot (al-

though this information has not been used to train the model). The assumed Gaussian process surrogate predicts the probability distribution of the identified acoustic source location conditional on the training data and though the mean true location does not match the exact source location (which is expected due to experimental errors and measurement noise), the true source location is contained within the 95% probability envelope as shown in Fig. 5(c).

Thus effectiveness of the methodology is demonstrated in terms of the accuracy of identifying the source of the AE. The cost of producing the gaussian process surrogate is incurred mostly in an offline training stage while in the online identification stage the effective location of the AE source is derived efficiently from the surrogate. This allows for implementation of the technique in real-time identification procedure within the performance constraints of the signal processing platforms.

The cross-correlation measures are transformed to the time-frequency domain using the wavelet transform, following Eqn. 15, and shown in Fig. 6. The wavelet

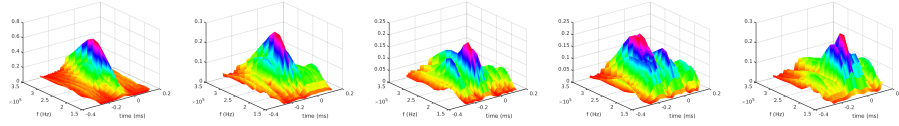


Fig. 6. The wavelet transformed surface of the cross-correlation measure between sensor pairs i.e. $\left| \mathbf{r}_{ij}^{s,\Delta} \right|$ where $i = 1$ and $j = 1, \dots, 5$

transformed surfaces show the relative positions in the time-scale of the point of arrival of the correlated wavemodes at the individual sensors in addition to the degree of correlation of the individual wavemodes (as amplitude along the z-axis). These are used as additional features (in addition to the normalized cross-correlated metrics given in Fig. 3).

Using the features shown in the cross-correlation spectrum and features from the wavelet transformed surfaces given in Fig. 6, the localization of the damage on the multilayered composite structure with H-N sources have been performed and shown in Fig. 7. This shows the location of five different AE sources using the isoprobability contours while the actual location is shown with the red marker. The figure superimposes identification performed for multiple AE source signals and the isoprobability contours signify the probability of source location based on the information collected from the 5 sensors distributed on the surface of the composite panels. The results show a good accuracy of AE source localization based on the identification features derived from the cross-correlation metrics and does not rely on any assumed shape/geometry or ad hoc regularization conditions applied for identification. Thus this has the potential to be applied in industrial scenarios as a robust identification technology for acoustic source location. The proposed methodology would also be applicable for structures operating under service loads (with the energy content concentrated in the

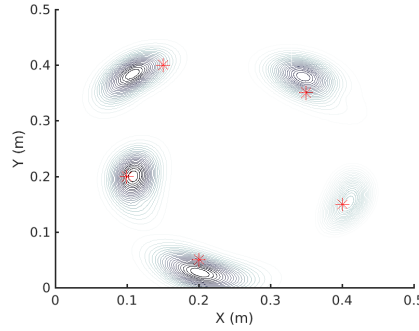


Fig. 7. The accuracy of localization of damage at multiple locations. The true AE location sources are marked with * while the isoprobability contours describe the possible location of the AE sources.

low frequency regime) which is a fundamental advantage with AE-based SHM techniques.

Another set of tests were performed on the composite panel, when it was subject to compressive fatigue loading cycles between 0.5 – 5 kN (peak-to-peak) for 1000 cycles at a frequency of 1 Hz. The distributed sensory network ‘listened’ to damages developing in the composite structure and waveforms of the signal that were collected has been utilized for a) source localization and b) characterization of the nature of damage. The source localization was performed following the same procedure reported earlier in this section for H-N sources based on the wavelet transformed features of the full correlation measure of the signals at the sensors. The damage classification was performed using the k -nearest neighbours (KNN) algorithm. The KNN algorithm is a supervised, non-parametric and instance-based method used for classification.

A training dataset comprising of a mixture of H-N sources and fatigue test data was used to train a KNN model. The training data set thus comprised of signal features X_f with a categorical variable c pairs as $\mathcal{D} = \{(X_f, c)_i, i = 1, \dots, n_{train}\}$. The categorical variable c takes binary values of $[0, 1]$ which correspond to the two different class of tests - the H-N source and the fatigue loading. The Mahalanobis distance was used to determine the distance in the feature space for the test data and a probabilistic estimate of the class variable associated with each test data point was determined. It has been observed that a validation accuracy of 97% was obtained for the aggregate of test data using the KNN algorithm. Thus the classification algorithm based on the signal features were successful in identifying the acoustic from H-N test data to that of fatigue loading.

The strength of the KNN algorithm in accurate classification of the test type shows the potential of the used signal features for identifying the nature of damage. The application of the proposed methodology in this paper to identify the generation of cracks in the composite panel under fatigue loading cycles

shows the potential of the proposed method for an automated probabilistic identification framework to improve the accuracy and minimize the computational overhead associated with the damage detection in composites.

7 Conclusion and future work

The study demonstrates the effectiveness of the proposed correlation feature based identification of AE sources in applications of non-destructive testing. The effectiveness of the method stems from using direct evaluations of normalized autocorrelation functions which effectively reduces the impact of signal noise and captures the essential correlation information between the signals recorded with the distributed sensor network. A reduced set of basis vectors which compress the correlation information reduces the memory and computational overhead associated with the method. A Gaussian process surrogate is fitted on the error surface over the spatial domain to explicitly consider the error associated with deterministic evaluators. The results demonstrate the effective of the method in predicting the source location with test data on the same panel.

The promises of the study extends to a number of interesting future investigations for extending and improving the proposed methodology to address additional challenges. Some of these are

- The wavelet transformed features of the correlation metrics can be used to derive the energy distribution in propagating modes from which the characterization of the damage type (based on categorical damage models would be accomplished) can be performed.
- The portability of the trained surrogate i.e. the ability of the trained model to predict AE sources in nominally similar test panels.
- A hierarchical probabilistic model of the AE source mapped to the correlation matrix, where identification stages can be compartmentalized to correspond to various levels of refinement.

Subsequent work would also focus on using physics based model of ultrasonic wave propagation in composite panels which would provide a means of identifying the nature of damage induced in the structure which has not been addressed in this study. This would allow us to go beyond the black-box input-output mapping techniques which is expected to significantly improve the performance of the identification algorithm and provide additional information pertaining to the degradation of health of operational structures.

References

1. Kundu, A., Adhikari, S., Friswell, M.I.: Stochastic finite elements of discretely parameterized random systems on domains with boundary uncertainty. *International Journal for Numerical Methods in Engineering* **100**(3), 183–221 (2014)
2. Miller, R., Carlos, M., Findlay, R., Godinez-Azcuaga, V., Rhodes, M., Shu, F., Wang, W.: *Acoustic Emission Source Location*, 3rd edn., pp. 121–146. ASNT (2005)

3. Eaton, M., Pullin, R., Holford, K.: Towards improved damage location using acoustic emission. *Journal of Mechanical Engineering Science, Proceedings of the Institution of Mechanical Engineers Part C* **226**(9), 2141–2153 (2012)
4. Kennedy, M.C., O'Hagan, A.: Bayesian calibration of computer models. *Journal of the Royal Statistical Society Series B-Statistical Methodology* **63**(3), 425–450 (2001)
5. Kundu, A., Matthies, H., Friswell, M.: Probabilistic optimization of engineering system with prescribed target design in a reduced parameter space. *Computer Methods in Applied Mechanics and Engineering* **337**, 281–304 (2018)
6. Miller, R., Anastasopoulos, A., Carlos, M., Demeski, R., Vallen, H., Walker, J.: *Acoustic Emission Signal Processing*, 3rd edn., pp. 147–180. ASNT (2005)
7. Ziola, S., Gorman, M.: Source location in thin plates using cross-correlation **90**, 2551–2556 (1991)
8. Aljets, D., Chong, A., Wilcox, S., Holford, K.: Acoustic emission source location in plate like structures using a closely arranged triangular sensor array. *Journal of Acoustic Emission* **28**, 85–98 (2010)
9. Hamstad, M., O'Gallagher, A., Gary, J.: A wavelet transform applied to acoustic emission signals: Part 2: Source location. *Journal of Acoustic Emission* **20**, 62–82 (2002)
10. Lokajicek, T., Klima, K.: A first arrival identification system of acoustic emission (ae) signals by means of a higher-order statistics approach. *Measurement Science and Technology* **17**, 2461–2466 (2006)
11. Akaike, H.: Markovian representation of stochastic processes and its application to the analysis of autoregressive moving average processes. *Annals of the Institute of Statistical Mathematics* **26**, 363–387 (1974)
12. Kurz, J.H., Grosse, C., Reinhardt, H.W.: Strategies for reliable automatic onset time picking of acoustic emission and of ultrasound signals in concrete. *Ultrasonics* **43**, 538–546 (2005)
13. Pearson, M., Eaton, M., Featherston, C., Pullin, R., Holford, K.: Improved acoustic emission source location during fatigue and impact events in metallic and composite structures pp. 1–18 (2016)
14. Kundu, T., Das, S., Jata, K.: Point of impact prediction in isotropic and anisotropic plates from the acoustic emission data. *Journal of the Acoustical Society of America* **122**(4), 2057–2066 (2007)
15. Kundu, T., Das, S., Martin, S., Jata, K.: Locating point of impact in anisotropic fibre reinforced composite plates. *Ultrasonics* **48**, 193–201 (2008)
16. Ciampa, F., Meo, M.: A new algorithm for acoustic emission localisation and flexural group velocity determination in anisotropic structures. *Composite Part A* **41**(12), 1777–1786 (2010)
17. Baxter, M., Pullin, R., Holford, K., Evans, S.: Delta T source location for acoustic emission. *Mechanical Systems and Signal Processing* **21**(3), 1512–1520 (2007)
18. Eaton, M.J., Pullin, R., Holford, K.: Acoustic emission source location in composite materials using delta T mapping. *Composites Part A* **43**(6), 856–863 (2012)
19. Al-Jumaili, S., Pearson, M., Holford, K., Eaton, M., Pullin, R.: Acoustic emission source location in complex structures using full automatic delta T mapping technique **72–73**, 513–524 (2016)
20. Scholey, J., Wilcox, P., Wisnom, M., Friswell, M.: A practical technique for quantifying the performance of acoustic emission systems on plate-like structures. *Ultrasonics* **49**, 538–548 (2009)
21. Hsu, N.N., Breckenridge, F.R.: Characterization and calibration of acoustic emission sensors. *Materials Evaluation* **39**(1), 60–68 (1981)

22. ASTM.: A standard guide for determining the reproducibility of acoustic emission sensor response. American Society for Testing and Materials, **E976** (2010)
23. Schumacher, T., Straub, D., Higgins, C.: Toward a probabilistic acoustic emission source location algorithm: A Bayesian approach **331** (2012)
24. Zarate, B., Pollock, A., Momeni, S., Ley, O.: Structural health monitoring of liquid-filled tanks: a Bayesian approach for location of acoustic emission sources **24** (2014)
25. Kundu, A., Eaton, M.J., Al-Jumali, S., Sikdar, S., Pullin, R.: Acoustic emission based damage localization in composites structures using Bayesian identification. *Journal of Physics: Conference Series* **842** (2017)
26. Kundu, A., Sikdar, S., Eaton, M., Navaratne, R.: Probabilistic method for damage identification in multi-layered composite structures. *Proceedings of the 9th European Workshop on Structural Health Monitoring* (2018)
27. Kundu, A., DiazDelaO, F., Adhikari, S., Friswell, M.: A hybrid spectral and meta-modeling approach for the stochastic finite element analysis of structural dynamic systems. *Computer Methods in Applied Mechanics and Engineering* **270**(0), 201 – 219 (2014)
28. Oakley, J.E., O’Hagan, A.: Probabilistic sensitivity analysis of complex models: a Bayesian approach. *Journal of the Royal Statistical Society B* **66**(3), 751–769 (2004)
29. Keane, A., Nair, P.: *Computational Approaches for Aerospace Design*. John Wiley & Sons, Chichester, UK (2005)
30. Sacks, J., Welch, W., Mitchell, T., Wynn, H.: Design and analysis of computer experiments. *Statistical Science* **4**(4), 409–435 (1989)
31. O’Hagan, A.: *Bayesian Statistics 4*, chap. Some Bayesian numerical analysis, pp. 345–363. Oxford University Press, Cambridge, UK (1992)
32. Haylock, R., O’Hagan, A.: *Bayesian Statistics 5*, chap. On inference for outputs of computationally expensive algorithms with uncertainty on the inputs. Oxford University Press, Oxford, UK (1996)
33. Oakley, J.: Eliciting Gaussian process priors for complex computer codes. *The Statistician* **51**(1), 81–97 (2002)
34. Rougier, J.: Probabilistic inference for future climate using an ensemble of climate model evaluations. *Climatic Change* **81**(3), 247–264 (2007)