Uncertainty Analysis

Graeme Manson[1], Keith Worden[1], •S. G. Pierce[2] and Thierry Denoeux[3]

[1] *Department of Mechanical Engineering, University of Sheffield, Sheffield, UK*
[2] *Department of Electronic and Electrical Engineering, University of Strathclyde, Glasgow, UK*
[3] *Université de Technologie de Compiègne, Compiègne, France*

1 INTRODUCTION

Future engineering will rely increasingly on computer simulations and modeling for a wide range of processes in the design-to-decommissioning life cycle—from design and prototyping, through production planning, manufacture, and in-service monitoring and ending with a safe decommission. This "virtual" approach usually offers clear financial and/or environmental advantages. However, a fundamental concern is that the modeling process must be robust to uncertainty: a high-fidelity model can become worthless if there is no understanding of how uncertainty is propagated through it. To elaborate on the concept of uncertainty propagation, consider a complex component incorporating inelastic materials, contacts, friction, etc. The model requires an initial specification of material properties, clearances, and friction coefficients. The response of the model will only be as accurate as these initial values, yet the values may be subject to large uncertainty; consider the difficulty in the mechanical characterization of biomaterials or viscoelastic polymers. In a worst-case analysis, the predictions of the model may prove highly sensitive to variations in the input parameters.

In 1999/2000, members of the Engineering Analysis group, amongst others, at Los Alamos National Laboratories (LANL), carried out arguably the most ambitious calculation in the (short) history of structural dynamics. Using one of the most powerful computers in the world at the time, the platform *Blue Mountain*, the computation essentially attempted

Encyclopedia of Structural Health Monitoring. Edited by Christian Boller, Fu-Kuo Chang and Yozo Fujino © 2008 John Wiley & Sons, Ltd. ISBN: 978-0-470-05822-0.

to quantify the propagation of uncertainty through a nonlinear finite element (FE) model of a weapon component under blast loading [1]. The calculation made use of 3968 processors from the available 6000 and used concurrently 4000 ABAQUS/Explicit licenses. The analysis took over 72 h and would have required 17.8 years of equivalent single-processor time.

The scale of this calculation is testimony to the importance that uncertainty quantification and propagation (UQP) is inevitably going to assume in the immediate future of structural dynamics and also in the long term. One of the reasons for the LANL interest in UQP is that current test bans in the field of nuclear weaponry forbid the type of experimentation which would usually be used to validate physical and computational models. Another compelling reason for UQP is the fact that the environmental and operational conditions for certain engineering projects are unknown and cannot be estimated with certainty. Alternatively, conditions can be estimated, but cannot be recreated exactly in validation experiments. Consider the design and manufacture of an in-orbit facility. In this case, aspects of the true environment, such as weightlessness and absence of drag, can be reproduced to an extent in terrestrial experiments. However, consider the design of an exploration module intended to operate on the surface of Venus. Such systems are extremely expensive and will often have only one chance to succeed. The performance of the Hubble telescope or the Mars Lander Beagle 2 is an object lesson here.

Another prime motivator for uncertainty analysis is the need for risk assessment in safety-critical systems. In fact, one might argue that the modern origins of the subject are here with projects like the Nuclear Reactor Safety Study [2]. One of the main regulators in the design of structural health monitoring (SHM) systems for say, a civil aircraft, will be the need to eliminate false assertions of damage as a result of environmental and operational variations. Such false positives and the ensuing "no fault found" inspections would substantially increase the cost of ownership of the aircraft. Even more critical is the possibility of a false negative as a result of environmental uncertainty, with potential loss of life being the result.

The purpose of this article is to introduce the issue of uncertainty analysis in general and also its relevance to SHM. In the next section, brief descriptions of the most popular of the many frameworks for uncertainty representation are given. Section 3 discusses the three main uncertainty-related problems of relevance to structural dynamics, namely, quantification, fusion, and propagation. In order to illustrate the ideas of the preceding two sections in a realistic scenario, two case studies conducted on an aerospace structure, namely the wing of a Gnat trainer aircraft, are provided. The first case study, in Section 4, considers the issue of attaining certification for artificial neural network (ANN) damage classifiers through the assessment of the network's robustness to uncertainty. This case study involves the propagation of intervals through the network structure and examines whether the networks, which would be considered as optimal, in the traditional sense, are also most robust to uncertainty. In Section 5, the second case study considers evidence-based classifiers as an alternative to probabilistic classifiers for the problem of damage location. The Dempster–Shafer (DS) theory is employed to construct neural network classifiers with the potential to admit ignorance, rather than misclassify. The section considers issues of propagation and fusion in an evidence-based framework and compares the performance of DS neural networks with their probabilistic counterparts. The article ends with some brief conclusions.

2 THEORIES OF UNCERTAINTY

There are many frameworks available through which uncertainty may be represented and this section examines some of the more common theories, considers their pros and cons, and looks for applications in the literature of structural dynamics. It is not the intention to give an exhaustive review of the literature and, in particular, the large body of work in the high-frequency/•SEA arena is ignored. Q5

2.1 Classical probability theory

The classical probability theory is well known. In fact, if the variations in a parameter are random, there is no better specification of the uncertainty than a probability distribution. However, in practice, this is often not available. Engineering analysis is routinely based on small samples and one might only be able to

estimate the low-order moments of a distribution—mean and variance—with any confidence. Arbitrarily imposing a known distribution shape e.g., Gaussian, on the basis of this information is perilous. In particular, the use of such central statistics may result in a distribution radically different in the tails from the true distribution. In risk analysis, where one is concerned with extreme events, the results of such a strategy will probably be meaningless. Another problem, in general, is that a specification of a problem will necessarily include a region of ignorance, and classical probability theory cannot accommodate this. In particular, a statement of the probability of an event automatically fixes the probability that it will not occur. The evidence for the occurrence of the event is essentially the same as the evidence for nonoccurrence.

2.2 Evidence theory

This can be regarded as an extension of the theory of probability (although this interpretation is contested). The main theory of this sort is the DS theory [3, 4], although generalizations exist [5]. Essentially, the single probability is replaced here with two quantities. The *belief, Be*, associated with an event is the sum of the evidence in favor of the event. The *plausibility, Pl*, is the complement of the evidence against the occurrence of the event. By the fact that these quantities can be different, the DS theory accommodates the idea of ignorance. The DS theory thus replaces the single probability with an interval [*Pl, Be*], and these quantities are sometimes called *lower and upper probabilities. Be* can be regarded as the best-case estimate of the probability of an event, corresponding to the case when all the missing evidence (ignorance) turns out to be in support of the event. Applications in engineering of the DS theory are rather few and far between, [6] cites 12 application references in total, but interestingly exclude [7, 8] and those that specifically consider problems of SHM.

2.3 Possibility theory

Similar to the DS theory, the possibility theory makes use of two complementary uncertainty measures—possibility and necessity [9]. Essentially, some proposition e is mapped into the interval [0,1], which may be divided into the three intervals: necessity $N(e)$, necessity of the contrary proposition $N(e')$, and ignorance $\theta(e)$. Possibility of the proposition is given by $P(e) = N(e) + \theta(e) = 1 - N(e')$. Applications of the possibility theory are even rarer than applications of DS, [6] cites only three. In some ways, the possibility theory can be interpreted in terms of fuzzy sets [10].

2.4 Fuzzy logic

This is one of the elder statesmen of contenders with the probability theory [11]. This extends the classical probability theory by relaxing one of the fundamental set-theoretic properties on which it is based. In the classical set theory, an element x is either a member of a set A or a member of its complement A^c. In fuzzy set theory, x may be associated (with given weight) with a number of different sets. Fuzzy logic encodes uncertainty by associating linguistic descriptors with a variable x like *large* or *small*. x may be a member of both sets, large and small, but it is associated to each by a membership function that mediates the likelihood of its membership. There are analogs of all the basic mathematical operations for such fuzzy variables (often based on interval arithmetic) and it is possible to construct fuzzy analogs of many *crisp* theories. The most intensive use of fuzzy logic in engineering is associated with control, although the interesting studies [12–14] construct a fuzzy version of modal analysis. The same authors have also formulated a fuzzy version of FE analysis [15].

2.5 Interval methods

Rather simply stated, the interval method replaces *crisp* numbers and variables with intervals $[\underline{x}, \overline{x}]$ [16]. Uncertainty in a parameter is encoded in the statement that it lies somewhere between two given bounds $\underline{x}$ and $\overline{x}$ with certainty. Note that this information could be incorporated in a probabilistic analysis by giving a distribution for the parameter on the interval, e.g., a uniform distribution. However, interval analysis makes no use of such additional

assumptions. There are again analogs of all the expected arithmetical operations, which thus allow the propagation of interval quantities through various types of models. There are numerous uses of interval arithmetic documented in the literature. In terms of structural dynamics, the previously cited study on fuzzy modal analysis, [13], made extensive use of interval arithmetic. In [17], there is an independent attempt to formulate a fuzzy FE analysis in terms of interval arithmetic. The paper [18] is of interest in that it considers a problem of damage detection. The problem with interval arithmetic is that it is conservative in nature and that the bounds on the calculations expand considerably in practical computations. An attempt to improve on this behavior is under development in the form of *affine arithmetic* [19]. An application to the structural eigenvalue problem can be found in [20].

2.6 Convex models

In a sense, the convex models can be regarded as a generalization of the interval concept of uncertainty (although they are more than this). A given parameter p is associated with a convex set, which may be said to contain the parameter. For example, an *ellipsoidal-bound* convex model takes the form

$$\Im(\alpha, p) = \{p : (p - \overline{p})^T W (p - \overline{p}) < \alpha\} \quad (1)$$

where W specifies the axes of the ellipsoid containing the data and $\overline{p}$ is the mean of the data set. This is a convex set. This approach to uncertainty was pioneered by Ben-Haim and has been applied by him and coworkers in numerous contexts [21]. It is possible to prove numerous theorems such as: if the input to a linear system is a convex model, then so is

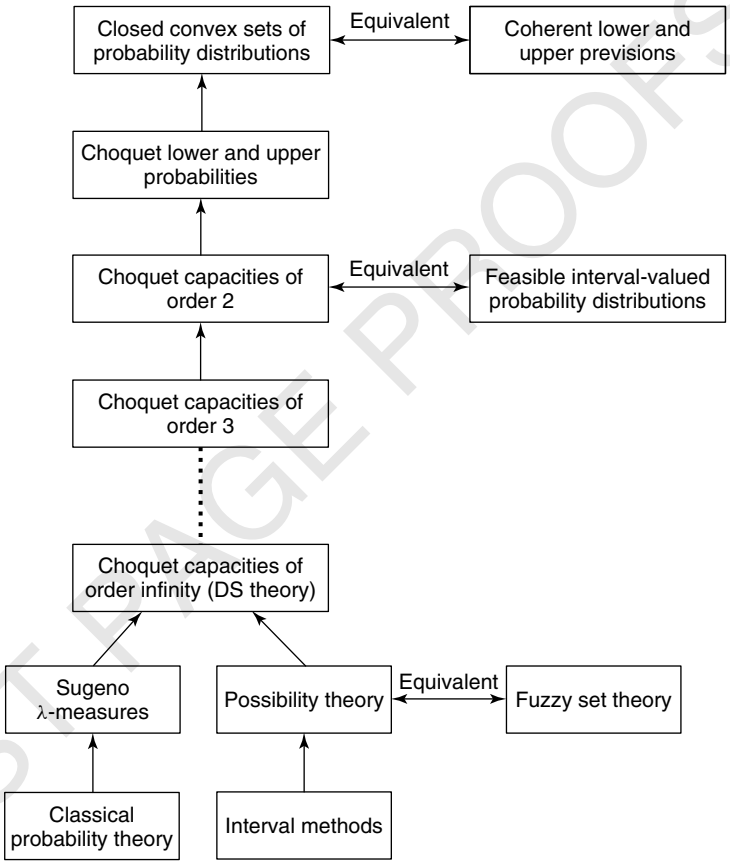

Figure 1. Klir and Smith's hierarchy of theories of uncertainty.

the output. This property does not hold for a nonlinear system, but by relaxing the constraints one obtains an *information-gap* model and recovers the property of invariance under a nonlinear operator [22]. An interesting formulation of an information-gap model on a problem of structural dynamics can be found in [23].

2.7 Klir and Smith's hierarchy of uncertainty theories

In a sense, the approaches described above are lower-level approaches to modeling uncertainty. In the survey [24], Klir and Smith identify a hierarchy of theories of uncertainty as shown in Figure 1.

As one moves up the diagram, one encounters more general representations of uncertainty. In practice, none of the theories above the DS theory have so far been applied to engineering problems; however, they have potential for the future. All the theories above the classical probability theory are based on a more general form of measure—*capacity* [25]. The classical probability theory has, for disjoint events A and B, that

$$p(A \cup B) = p(A) + p(B) \tag{2}$$

i.e., the measure p is *additive*. Choquet's capacities g allow

$$g(A \cup B) >< g(A) + g(B) \tag{3}$$

i.e., they account for the possibility of mutual cooperation or inhibition between events. As discussed above, this freedom has so far been neglected in terms of engineering applications.

3 QUANTIFICATION, FUSION, AND PROPAGATION OF UNCERTAINTY

In terms of uncertainty, structural dynamics has arguably three main concerns, which are discussed briefly in this section.

3.1 Quantification

Given a parameter a, associate with this a meaningful uncertainty measure with the framework of a given uncertainty model. For example, assign a probability in the classical probability theory or a basic probability assignment (BPA) in the DS theory. The process of quantification may involve several steps, each refining the measure. For example, in [26], the following process is described within the context of the Bayesian probability theory. A prior distribution is defined, for example, for a given model parameter. If it exists, expert opinion data is used to update/refine the priors to obtain a posterior distribution for the parameter. This posterior then becomes the prior for stage 2 whereby the results of computer simulation are used to update. Finally, the posterior from stage 2 becomes the prior for stage 3 where experimental data are used to update. The multistage process may also be thought of as a precursor to fusion.

3.2 Fusion

Given a parameter a, which has a set of uncertainty measures $m(a)_i$ associated with different uncertainty models U_i, how does one refine each measure in the light of the others? For example, if one is given a classical probability 0.75 for an event, and lower and upper probabilities [0.6, 1.0] from the DS theory, how can one refine each measure given the other? Associated with this problem is that of *normalization* i.e., which uncertainty measure in the convex model theory corresponds to a given probability or BPA? Or is it even possible to make a quantitative comparison? This is largely an open problem, although some preliminary results are available; [27] sketches a relationship between the fuzzy set theory and the classical probability theory.

3.3 Propagation

If a parameter a, with a given uncertainty measure $m(a)$, forms the input to a given physical process, which uncertainty measure should be associated with the output, or a given aspect of the output? As an example, consider the discussion in [28] a control problem. The aim is to establish, given a system with uncertain parameters, the probability that the derived

controller will be stable. An associated problem here is the question of *sensitivity*, i.e., which parameters, with their associated uncertainties, are the main cause of variation in the model/system response.

4 CASE STUDY 1: ASSESSMENT OF NEURAL NETWORK ROBUSTNESS FOR DAMAGE CLASSIFICATION

The first case study considers the application of ANN classifiers to the interpretation of data for the purposes of damage location. ANN classifiers have found diverse applications in many fields including aircraft wing damage detection [29, 30]. Conventional network training can be viewed within a framework of the global optimization (minimization) of the error function between the network output prediction and the true target data. There exist a number of strategies to locate this minimum, the most common being variants of gradient descent such as scaled conjugate gradients [31, 32]. These techniques make use of the local gradient information of the error surface to ascertain the optimum search direction but are susceptible to the danger of locking a solution into a local minimum rather than locating the true global minimum. A number of strategies have been devised to counter this problem [31–33], and new techniques of searching the error function space are currently active areas of research, for example, using genetic algorithms (GAs) [32].

An additional complication to network performance lies in the capability of a network to generalize its classification performance to previously unseen data. There exists the widely recognized problem of overtraining that can occur such that a network starts to learn the noise present in data rather than the underlying data structure. The use of cross-validation and early stopping [33] using an independent validation data set are often used as termination criteria for training to help avoid this problem.

The problems of network overtraining and lack of generalization are central to understanding the inertia to the practical application of ANNs, especially to safety-critical applications. If, for example, one envisages an ANN classifier being used to assess the condition of a major structural airframe component, it is imperative that the performance of the classifier to the most diverse range of inputs is well understood. Poor classifier performance could result in catastrophic failure and possible loss of life. Although a range of techniques have been developed for output confidence interval predications [32, 34–37], they all adopt a probabilistic standpoint and therefore suffer from the common drawback that since the probability distributions are usually estimated from the low-order moments of the data (typically mean and standard deviation), there is often no representation associated with the extremes of the distributions. Unfortunately, it is often the extreme events of the data that are likely to be associated with the unpredictable failure events of greatest interest.

This case study comprises a novel nonprobabilistic approach applied to predicting extreme network outputs in the presence of uncertainty in the input data. This technique is based on the theory of convex models and information-gap uncertainty as pioneered by Ben-Haim [38–41]. Interval-based [42] techniques are extended to investigate the response of a simple multilayer perceptron (MLP) network used for a classification problem to locate damage sites on the wing of a Gnat trainer aircraft. A comparison with conventional network training based on cross-validation is presented. It is shown that the use of interval-based network propagation allows a new criterion for network selection to be established. This technique allows the identification of an ANN classifier, which is intrinsically more robust to noise on the input data than network solutions obtained by conventional maximum likelihood training. Furthermore, by virtue of the conservative nature of interval sets, the reliance on probabilistic-based estimates of confidence bounds on network predictions is obviated. The interval-based worst-case error predictions represent an inclusive bounded solution set given a specified degree of input noise to the classifier.

4.1 Experimental data acquisition and signal processing

The work concerns an SHM strategy to the problem of damage location on an aircraft (Gnat) wing located at •DSTL Farnborough, as shown in Figure 2. Q6
The wing was instrumented with an array of 12 accelerometers to measure the response to forced

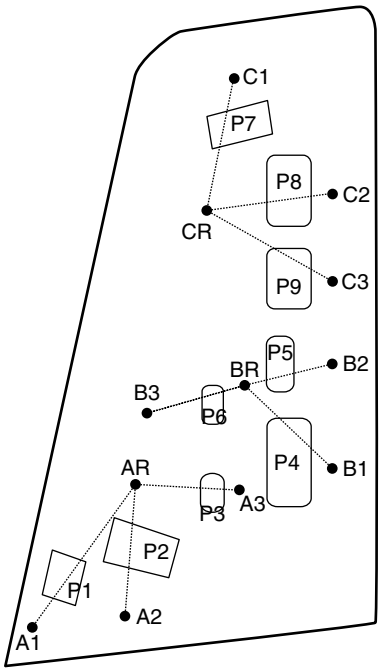

Figure 2. Detail of Gnat wing, showing position of sensors and removable panels.

vibration induced from a shaker mounted to the underside of the wing. The wing had a series of nine removable panels (P1–P9), which could be removed and replaced to provide a reproducible and reversible representation of changing conditions on the wing structure. In this fashion, damage represented by a local change in stiffness properties could be simulated.

Data was collected from all 12 accelerometers for a variety of undamaged (normal condition with all panels in place) and simulated damage data. Note from Figure 2 that the accelerometers were positioned in three distinct groupings (A, B, C) across the plate. Rather than record the individual acceleration responses, the experiment was configured to record the ratio of measured accelerations between transducer pairs AR/A1, AR/A2, AR/A3, etc. in such a way that the transmissibilities between transducer pairs formed the base measurement [29, 30]. In this fashion, there were a total of nine measurement variables recorded. Figure 3 illustrates a typical raw transmissibility measurement with 1024 spectral lines recorded with 1 Hz width in the frequency range 1024–2048 Hz.

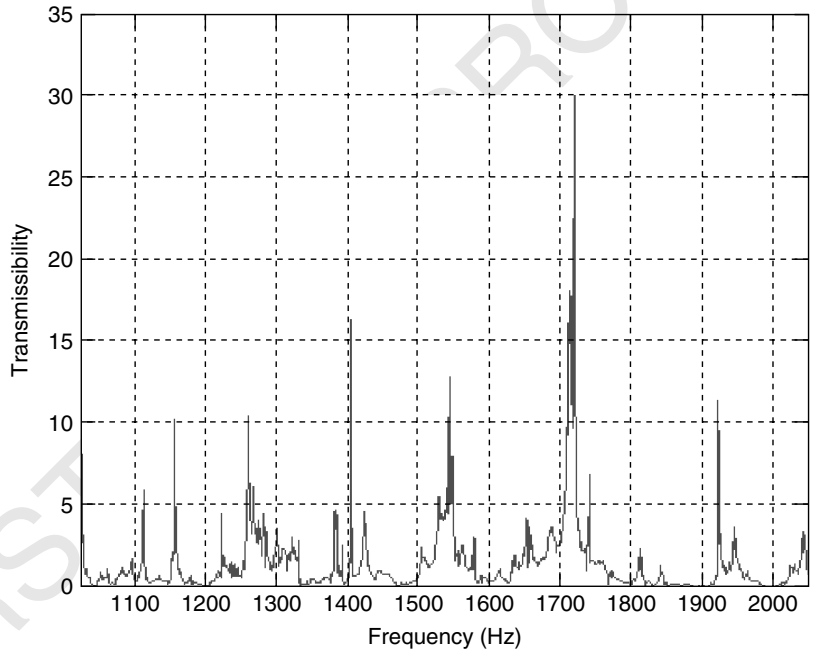

Figure 3. Example of a transmissibility plot spectrum.

By systematic removal of panels P1–P9, the effect of simulated damage on the spectral response of the transmissibility functions could be observed. For each panel removed, 100 individual measurements were recorded for each of the nine separate transmissibilities (labeled T1 through T9) corresponding to removal of panels P1–P9. Two complete runs through the set were completed making a total of 18 runs each with 9 transmissibilities each with 100 measurements; giving a total of 16 200 results. Additionally, 7 normal condition (undamaged) cases were recorded for each of the 9 transmissibilities, again taking 100 individual measurements for each, thus giving a total of 6300 results representing normal condition.

The data was inspected manually to select features (spectral line windows) from the T spectra (transmissibility spectra) that corresponded to particular damage events. This problem was simplified by using a nonlocal argument with respect to the transducer groups [29, 30]. For removal of panels P1–P3, only the T spectra from T1, T2, T3 (corresponding to the accelerometers A1–A3 and AR) were considered relevant, and the spectra from transducer groups B and C were ignored. Similarly, for the removal of panels P4–P6, only measurements from the B set of transducers were considered and likewise only measurements from the C set of transducers for the removal of P7–P9.

From a total of 77 individual features [29, 30], a single feature was selected (by inspection) to correspond to the removal of a single panel. In this fashion, the feature set was reduced to nine individual features (F1–F9). For each individual feature, the data comprised 700 normal condition measurements, and 1800 test measurements. An outlier (novelty) analysis was then performed [43] by computing the Mahalanobis squared distance of the data points with respect to the normal condition data.

$$\Delta^2 = (x - \mu)^T \mathbf{\Sigma} \Sigma^{-1} (x - \mu) \tag{4}$$

where Δ is the Mahalanobis distance, x is the data, μ is the mean of the normal condition data, and $\mathbf{\Sigma}$ is the covariance matrix of the data.

By performing the novelty analysis for each of the nine features, a data matrix of size [1800 by 9] was obtained. This was divided into three equal parts to form separate *training, validation*, and *test* data sets for subsequent network evaluation. Finally, the data sets were logarithmically compressed and normalized to -1 to $+1$ before presentation to the network.

4.2 Network topology and training

The MLP network implementation and training was undertaken in MATLAB™ using the NETLAB toolbox developed by Nabney [32]. The data was presented to a series of MLP networks with different numbers of hidden nodes. Each network had nine input nodes corresponding to the features F1–F9, and nine output nodes corresponding to the classes P1–P9 (Figure 4). The input values were x_i, $i = 1, \ldots, d$. The outputs from the second layer were given by

$$a_k^{(2)} = \sum_{j=1}^{M} w_{kj}^{(2)} \tanh \left[\sum_{i=1}^{d} w_{ji}^{(1)} x_i + b_j^{(1)} \right] + b_k^{(2)} \quad \text{for } j = 1, \ldots, M \quad \text{and} \quad k = 1, \ldots, C \tag{5}$$

where w is the weight matrix, b is the bias matrix, M is number of input nodes, and C is number of output nodes.

The network output was given by transformation of the second layer activations by the output activation function. Since there were a series of C independent output classes, it was appropriate to utilize the *softmax* function [32]:

$$y_k = \frac{\exp[a_k^{(2)}]}{\sum_{k'} \exp[a_{k'}^{(2)}]} \tag{6}$$

The choice of the softmax activation function ensured that the outputs always summed to unity, and thus could be directly interpreted as class conditional probability values.

The number of hidden nodes in the second layer was varied between 1 and 15 hidden units. Each individual network structure was trained with 100 independent training sessions starting at differently randomly chosen points on the error surface so that a total of 1500 independent networks were evaluated. Up to 1000 iterations of a scaled conjugate gradient optimization were implemented within a maximum likelihood training framework using a small hyperparameter $\alpha = 0.001$ to control weight decay.

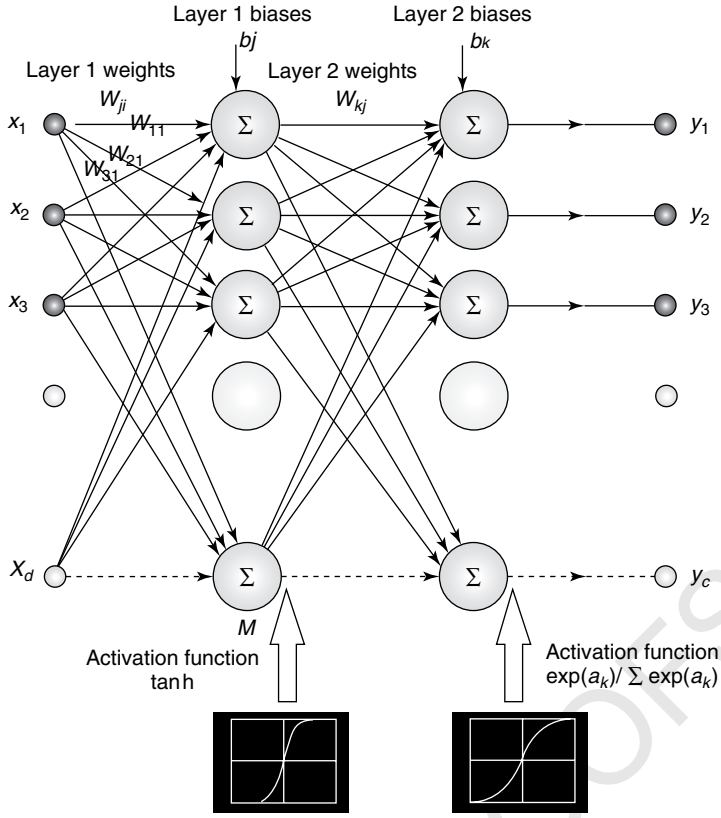

Figure 4. MLP network structure●.

Q2

4.3 Conventionally trained network results

Traditional network selection was implemented by dividing the data into three equal portions designated the training, validation, and test data sets. The training data was used to train the networks; the best network was then selected from the 1500 possibilities by finding the best classification rate on the validation data set. Finally, the performance of this selected best network was assessed using the test data set. It was found that a network with eight or nine hidden nodes produced an excellent classification performance for a relatively compact network structure.

However, considering that an eight-hidden-node network structure has 161 independent weight and bias components and the total number of training examples was 594 (66 examples of each class), it is very likely that the network structure with eight hidden nodes was likely to possess poor generalization performance. It was for this reason that the maximum network size considered in the present analysis was four hidden nodes, which have 22 independent weight and bias components. This choice of network provides reasonable classification performance on the validation data, whilst maintaining a relatively simple network structure. Using a cross-validation [33] approach to network selection, the best-performing network (with four hidden nodes) from the validation set was selected. This network gave a classification rate of 92.9% when applied to the test data set.

4.4 Interval-based classification networks and network robustness

Having established the network performance to *crisp* (i.e., single-valued) input data, the next step was

to devise a method to investigate the sensitivity of the classification performance of the network to fluctuations in its input data. Quantification of this performance would allow an estimate of the network robustness to be evaluated. Perhaps the most obvious way of performing this analysis would be to use a Monte Carlo approach, randomizing the input data (within certain predefined bounds) and monitoring the associated changes in the output classification performance. This technique has a significant drawback, especially when applied to nonlinear MLP networks, in that it is impossible to be sure of mapping all possible combinations of variation in input space to output space unless an unfeasibly large number of sample points are used. Since interest lies in understanding the worst possible performance of the classifier in the presence of input data uncertainty, a Monte Carlo approach would therefore not provide certainty of behavior under all possible input data conditions. Similarly, the other techniques discussed in the introduction have a similar flaw in that they generate a probabilistic view of the likelihood of the classifier performance with respect to input data fluctuation. It was to circumvent this problem that the input data set was redefined as a series of interval number inputs. Interval numbers [42] occupy a bounded range of the number line, and can be defined as an ordered pair of real numbers $[a, b]$ with $a < b$ such that

$$[a, b] = \{x | a \leq x \leq b\} \tag{7}$$

Interval numbers have specific rules for the standard arithmetic operations of addition, subtraction, multiplication, and division [40].

The MATLAB™ compatible toolbox INTLAB [44] was used to implement the interval calculations required to calculate forward propagation through the MLP networks. This toolbox incorporated a rigorous approach to rounding, which was critical when using finite precision calculations in order to preserve the true conservative interval bounds. Formally, with each network, we associated an input set $I(\beta)$ composed of all possible inputs to the network, the size of the uncertainty was governed by the β parameter. Having defined the input set $I(\beta)$, the output response set $R(\beta)$ of all network outputs was computed. It was then possible to quantify the network reliability in terms of how large a β parameter could be tolerated before a point in the failure set (defined by choosing an appropriate threshold for a performance governing parameter, for example, the worst-case error) was just reached; at this point, β attained a critical value β_{CR}. A large value of β_{CR} was desirable as the network would then be more robust [41].

Each input value x_i of the test data set was intervalized by a parameter β such that

$$[x_{ia}, x_{ib}] = [(x_i - \beta), (x_i + \beta)] \tag{8}$$

Propagation of interval sets through a crisp-valued ANN weight matrix gives rise to interval number outputs. For a regression problem, this is manifest as a set of upper and lower bounds around the true output prediction. However, for a classification problem, it is necessary to introduce some new definitions for the concepts of the confusion matrix and classification rate to be valid.

4.4.1 Defining interval classification rates

For crisp-valued outputs, the designation for a correct classification is if the class with the highest class conditional probability output (the winning class) belongs to the correct target class. If this is not the case, then a misclassification is assigned. The overall classification rate is just the percentage of classifications to total number of data presentations. For an intervalized network output with sufficiently small interval bounds, it would be expected that the classification rate would be the same as in the crisp data situation. The class with the highest bound output remains the winning class (correct or incorrect), and its lower bound is greater than any of the other classes upper bounds. However, as the interval size increases, a point is reached where one (or more) of the losing class upper bounds becomes equal to or greater than the lower bound of the winning class. (The lower bound of the winning class is defined as the *threshold value* for interval-based classification.) At this critical point, it becomes impossible to distinguish between the two (or more) classes, and either (or any) of the classes could be the winner. The best-case classification rate is defined as the percentage of correct classifications, regardless of whether these classifications were unambiguous (only one class identified) or ambiguous (more than one class identified). The worst-case classification rate is

defined as the percentage of unambiguous, correct classifications. In keeping with the framework of Ben-Haim [39, 41], opportunity is defined as the best classification rate minus the crisp classification rate. It is clear that, as the interval size increases, the worst-case classification rate decreases, whilst the best-case classification rate and the opportunity increase.

4.4.2 Interval propagation through conventionally trained network

The network with four hidden nodes, which was previously chosen via the cross-validation approach, was then subjected to the propagation of intervalized data in order to investigate its robustness to data uncertainty. The behavior of this network is shown in Figure 5.

This is a typical information-gap style plot [40, 41], where the opportunity is a measure of how much performance headroom is available to the system if we are prepared to tolerate the presence of uncertainty in the input data. However, for safety-critical systems, we are more likely to be interested in the worst-case classification rate. From such a figure, it is possible to decide on a minimum acceptable classification rate (for example 80%) and then infer the corresponding interval size, which in turn relates to the spread in the input data. For the case of Figure 5, the 80% minimum worst-case classification rate corresponds to an uncertainty in the input data of $\beta_{\mathrm{CR}} = 0.01$. Since the interval output set is conservative, we can then guarantee that, if the input set remains within the bound specified by β_{CR}, the output classification rate cannot fall below the minimum specified rate of 80%. In this fashion, the interval forward propagation routine can be used to assess the robustness of an individual network to uncertainty in the input data [45].

4.5 Interval-based network selection

Having established the basis for the definition of classification rates, worst-case error, opportunity, and best-case error applied to interval outputs from an individual network [28], the technique was extended to investigate propagation through multiple networks to appraise the capability of using interval results as a basis for network selection. The same networks trained using the conventional maximum likelihood approach detailed above were investigated. For each particular value of the number of hidden nodes (1–15), the interval forward propagation was evaluated for all 100 individual networks. The range of interval outputs for the worst-case error, opportunity,

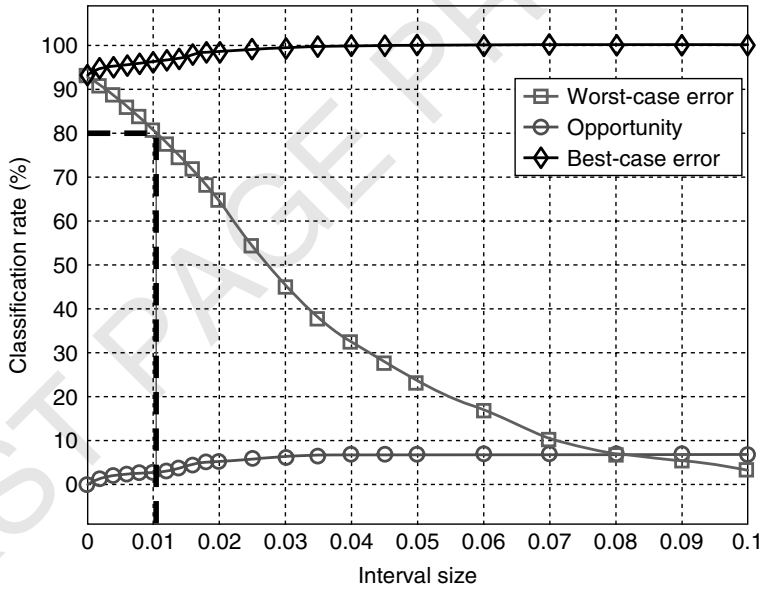

Figure 5. Classification rate as a function of interval size for four hidden node network chosen using cross-validation.

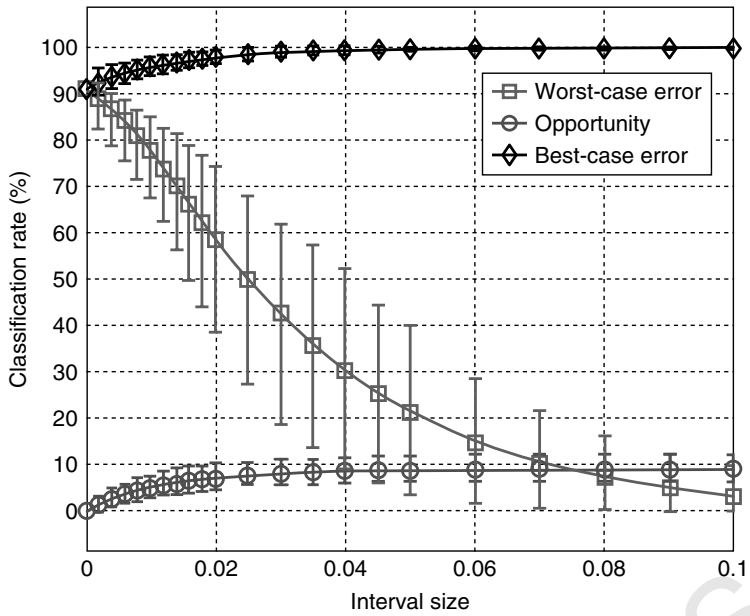

Figure 6. Interval output classification rates as function of interval size for all 100 of the 4 hidden node networks.

and best-case error were then calculated. The results for all 100 of the 4 hidden node networks are shown in Figure 6, where the markers indicate the mean values and the error bars show the range (i.e., the maximum and minimum values) of the three functions.

The most interesting feature of Figure 6 is the large spread in the values of the worst-case error function. For example, at an interval size of 0.02, the worst-case error falls between classification rates of 38.7 and 74.6%. The best-performing network using conventional cross-validation on crisp data, whilst not the worst-performing network, was certainly not the best. The best-performing network in terms of highest worst classification rate varied, depending upon the interval size but a few networks performed consistently well over the entire interval range.

In general, to select the best network in an interval tolerant sense, it is required to set a sensible upper limit on the input uncertainty (the interval size. The best network would then be the one giving the highest worst-case error at this uncertainty level. It is imperative to then check to ensure that this network also provides good performance at lower interval values. In this fashion, when considering interval input data, it is possible to determine a more appropriate choice of network selection criteria than is available using the conventional maximum likelihood training applied to crisp input data.

The merit of the interval-based forward propagation technique lies in two distinct areas. Firstly, it allows the definition of a critical size of uncertainty in the input data set that guarantees that the output classification performance does not fall below a predefined level. This is useful to apply to a single network structure (which could be obtained from any general training technique) to evaluate the robustness of that particular network to uncertainty or noise in the input data. Secondly, and more importantly, an extension of the first technique can be used as an alternative criterion for network selection. It was demonstrated that the worst-case interval output classification rates from conventional MLP networks can be used to find network structures with significantly improved classification performances over their conventional crisp-output counterparts in the presence of input uncertainty or noise. In general, it seems that prior knowledge of the maximum size of the input uncertainty is required to select the optimum network structure.

5 CASE STUDY 2: EVIDENCE-BASED DAMAGE CLASSIFICATION FOR AN AIRCRAFT STRUCTURE

The recent past has seen considerable use of machine learning techniques for SHM. The basic idea of the approach is to use data measured from undamaged and damaged structures in order to train a learning machine to assign a condition label to previously unseen data. The simplest problem of SHM is arguably *damage detection*. This is most easily carried out in the machine-learning context by using a novelty detector [46]. Novelty detection involves the construction of a model of the normal condition of a system or structure, which can then be used in a hypothesis test on unseen data to establish whether the new data corresponds to normal condition or not. The advantage of the novelty detection approach is that it can be carried out using unsupervised learning, i.e., with only samples of undamaged data. If a more detailed diagnosis of a system is required, e.g., if it is necessary to specify the type or location of damage in a structure, this can still be done using machine-learning methods. For higher levels of diagnostics, algorithms based on classification or regression are applicable; however, these must be applied in a supervised learning context and examples of data from both the undamaged and damaged conditions can be used [47].

The most popular classifiers for damage location and quantification so far have been those based on MLP neural networks [31] (although there is growing popularity for classifiers based on the concepts of statistical learning theory—like support vector machines [47]). Training of MLP networks as classifiers is usually accomplished by using the 1 *of M* strategy [31], which implicitly assumes a Bayesian probabilistic basis for the classification. Whilst the probability theory is only one (but arguably the most important) of a group of theories that can quantify and propagate uncertainty, other theories of uncertainty, perhaps with the exception of the fuzzy set theory, have been largely unexplored in the context of damage identification. The object of this article is to design a classifier for damage location based on the DS theory of evidence [3, 4]. The reason for exploring the possibilities of the DS theory is that it *extends* probability in a number of ways, which are potentially exploitable in an SHM context. The current case study is looking only to demonstrate that the method is competitive with the probability-based MLP approach on an experimental case study of an aircraft wing. The DS classifier here is also implemented using a neural network structure [48].

5.1 Dempster–Shafer reasoning

The DS theory is a means of decision fusion, which is formulated in terms of probability-like measures but extends the probability theory in a number of important respects. The basic idea of *belief* was introduced by Dempster in [3] and extended in Shafer's treatise [4].

The basic model is formulated in similar terms to probability. In the place of the sample space is the *frame of discernment* Θ, which is the set spanning the possible events for observation A_i, where $i = 1, \ldots, N$. •On the basis of sensor evidence, each Q7 event or union of events is assigned a degree of probability mass or *basic belief assignment* (*BBA*) m such that,

$$0 \leq m(A_i) \leq 1 \quad \forall A_i \subseteq \Theta \tag{9}$$

$$m(\phi) = 0 \tag{10}$$

$$\sum_{A_i \subseteq \Theta} m(A_i) = 1 \tag{11}$$

where φ is the empty set. The difference between this *evidential* theory and the probability theory is that the total probability mass need not be exhausted in the assignments to individual events. There is allowed to be a degree of *uncertainty*. This is sometimes denoted by a probability mass assignment to the *whole* frame of discernment $m(\Theta)$ or $m(A_1 \cup \ldots A_N)$.

The *belief* in an event B is denoted by $Be(B)$ and is defined by

$$Be(B) = \sum_{A_i \subseteq B} m(A_i) \tag{12}$$

and this is the total probability, which is committed to the support of the proposition that B has occurred.

The *doubt* in the proposition B is denoted by $Dou(B)$ and is defined by

$$Dou(B) = Be(\sim B) \tag{13}$$

i.e., the doubt is the total support for the negation of the proposition B (negation is denoted by a tilde).

One of the fundamental differences between the DS and the probability theory is that the belief and doubt *do not necessarily sum to unit* i.e., it is not certain that $B \cup B$ is true. This is illustrated diagrammatically in Figure 7.

The uncertainty Un in the proposition B is that portion of the probability mass which does not support B *or* its negation. If further evidence were provided, some of the uncertainty could move in support of B but the mass assigned to the doubt *cannot* move. This means that the possible belief in B is bounded above by the quantity $Be(B) + Un(B) = 1 - Dou(B)$ and this quantity is termed the *plausibility* of B and is denoted by $Pl(B)$. The plausibility can also be defined by

$$Pl(B) = \sum_{A_i \cap B \neq \phi} m(A_i) \tag{14}$$

A concrete example will be useful at this point.

Example *Consider a composite structure which may have sustained damage at one of two internal sites A and B which are indistinguishable. It is known that the only possible damage mechanism at site A is delamination (denoted D), but site B may fail by delamination or fiber fracture (denoted F) and the relative probabilities of the damage mechanisms are unknown. It is further known that failure at A is twice as likely as failure at B. What can one say about the likely damage type if a fault is found?*

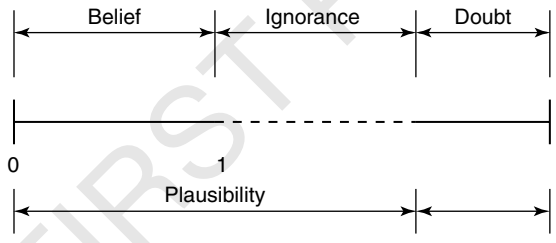

Figure 7. The Dempster–Shafer uncertainty interval.

First of all, if damage occurs at A, it is certainly by delamination and this forces the mass assignment,

$$m(D) = \frac{2}{3} \tag{15}$$

the remaining mass cannot be assigned with certainty, so it is assigned to the frame of discernment,

$$m(\Theta) = m(D \cup F) = \frac{1}{3} \tag{16}$$

The belief in the delamination is simply $Be(D) = 2/3$ *as this is the only basic mass assignment to B. There is no such assignment to F so the belief* $Be(F) = 0$. *The plausibility in D is given by*

$$Pl(D) = m(D) + m(D \cup F) = 1 \tag{17}$$

and the plausibility of F is similarly calculated as 1/3. *The uncertainty interval for D is* [2/3,1] *and that for F is* [0,1/3].

Note that it is not possible to use the probability theory here directly as the relative probabilities at site B are not known. It is possible to construct bounds on the probabilities though. Suppose delamination were impossible at site B, then the overall probability of delamination would be 2/3 *and this would be a lower bound. If delamination were certain at B, the overall probability would be* 1. *Note that these quantities are the belief and plausibility respectively. For this reason, the belief and plausibility are sometimes termed the lower and upper probabilities.*

The interpretation of some common instances of the uncertainty interval for a proposition B is as follows:

[0,0] B is certain.
[1,1] B is impossible.
[0.75,0.75] There is no uncertainty. B has a true probability of 0.75.
[0,1] There is total ignorance regarding B.
[0.25,1] B is plausible, there is no support for $\sim B$.
[0,0.75] $\sim B$ is plausible, there is no support for B.
[0.25,0.75] Both B and $\sim B$ are plausible.

All this suffices to establish terminology, to explain how to compute belief functions, and how to interpret the results. It does not provide a means of data

fusion—that requires the use of *Dempster's combination rule* [49].

Suppose that one has two sensors 1 and 2. BPAs are possible on the basis of either sensor alone, denoted m_1 and m_2. Belief functions Be_1 and Be_2 can be computed. Dempster's rule allows the calculation of an overall probability assignment m_+ and a corresponding overall belief function Be_+, where this *direct sum*,

$$Be_+ = Be_1 \oplus Be_2 \tag{18}$$

is induced by m_+. Suppose that sensor 1 makes assignments $m_1(A_i)$ to the proposition A_i (which can, and usually does, include the frame of discernment), and sensor 2 makes assignments $m_2(B_j)$, then Dempster's rule makes assignments as follows.

Consider a matrix with row entries labeled by i and column entries j, then the (i, j)th element of the matrix is an assignment of probability mass $m_1(A_i) \times m_2(B_j)$ to the proposition $A_i \cup B_j$.

This is best understood by an example:

Example *Suppose a classifier is required which can assign a damage type to data from a composite structure. The possible damage types are delamination D, fiber fracture F, matrix cracking M, or fiber pullout P. Two different classifiers are trained, A and B, which produce different probability mass assignments. Classifier A makes the assignments,*

$$m_A(D) = 0.25, \quad m_A(F \cup M) = 0.5, \quad m_A(\Theta) = 0.25 \tag{19}$$

and classifier B returns,

$$m_B(D \cup M) = 0.3, \quad m_B(D \cup F) = 0.4, \quad m_B(\Theta) = 0.3 \tag{20}$$

Dempsters rule induces a mass assignment matrix,

	$m_A(D)$	$m_A(F \cup M)$	$m_A(\Theta)$
$m_B(D \cup M)$	$0.25 \times 0.3 = 0.075$	$0.25 \times 0.4 = 0.1$	$0.25 \times 0.3 = 0.075$
$m_B(D \cup F)$	$0.5 \times 0.3 = 0.15$	$0.5 \times 0.4 = 0.2$	$0.5 \times 0.3 = 0.15$
$m_B(\Theta)$	$0.25 \times 0.3 = 0.075$	$0.25 \times 0.4 = 0.1$	$0.25 \times 0.3 = 0.075$

to the propositions,

	$m_A(D)$	$m_A(F \cup M)$	$m_A(\Theta)$
$m_B(D \cup M)$	D	M	$D \cup M$
$m_B(D \cup F)$	D	F	$D \cup F$
$m_B(\Theta)$	D	$F \cup M$	Θ

and the direct sum m_+ assigns support to the propositions D, M, F, $D \cup M$, $D \cup F$, $F \cup M$, and $\Theta = D \cup M \cup F \cup P$.

The belief in delamination is $Be_+(D) = 0.075 + 0.15 + 0.075 = 0.3$; similarly, $Be_+(M) = 0.1$, $Be_+(F) = 0.2$, and $Be_+(P) = 0$. The plausibility of delamination D is given by

$$Pl_+(D) = m_+(D) + m_+(D \cup M) + m_+(D \cup F) = 0.3 + 0.075 + 0.15 = 0.525 \tag{21}$$

similarly,

$$Pl_+(M) = 0.275, \; Pl_+(F) = 0.45 \; \text{and} \; Pl_+(P) = 0.075 \tag{22}$$

In summary, the uncertainty intervals for the fused belief function are

D	[0.3,0.525]
M	[0.1,0.275]
F	[0.2,0.45]
P	[0,0.075]

The most plausible diagnosis is clearly delamination.

In mathematical terms, Dempster's combination rule is expressed as

$$m_+(C) = \sum_{A_i \cap B_j = C} m_1(A_i) m_2(B_j) \tag{23}$$

and

$$Be_+(C) = \sum_{\substack{i,j \\ A_i \cap B_j = C}} m_1(A_i) m_2(B_j) \qquad (24)$$

Unfortunately, things are not quite as straightforward as this. Problems arise in using Dempster's rule if the intersection between supported propositions A_i and B_j is empty. In this circumstance, a nonzero mass assignment is made to the empty set ϕ and this contradicts the basic definition of the mass assignment, which demands that $m_+(\phi) = 0$. In order to preserve this rule, Dempster's rule *must* assign zero mass to nonoverlapping propositions. However, if this is the case, the probability mass is lost and the total mass assignment for m_+ is less than unity, contradicting another rule for probability numbers. A valid mass assignment is obtained by *rescaling* m_+ to take account of the lost mass. If the mass lost on nonoverlapping propositions totals k, the remaining mass assignments should be rescaled by a factor $K = 1/(1-k)$. The combination rule (23) is modified to

$$m_+(C) = K \sum_{A_i \cap B_j = C} m_1(A_i) m_2(B_j) \qquad (25)$$

Example *Consider the last example. Suppose the assignments made by sensor A were as before, but those of sensor B were now,*

$$m_B(D \cup M) = 0.3, \quad m_B(P \cup F) = 0.4,$$
$$m_B(\Theta) = 0.3 \qquad (26)$$

Dempster's rule gives the same assignments,

	$m_A(D)$	$m_A(F \cup M)$	$m_A(\Theta)$
$m_B(D \cup M)$	$0.25 \times 0.3 = 0.075$	$0.25 \times 0.4 = 0.1$	$0.25 \times 0.3 = 0.075$
$m_B(P \cup F)$	$0.5 \times 0.3 = 0.15$	$0.5 \times 0.4 = 0.2$	$0.5 \times 0.3 = 0.15$
$m_B(\Theta)$	$0.25 \times 0.3 = 0.075$	$0.25 \times 0.4 = 0.1$	$0.25 \times 0.3 = 0.075$

but this time to the propositions,

	$m_A(D)$	$m_A(F \cup M)$	$m_A(\Theta)$
$m_B(D \cup M)$	D	M	$D \cup M$
$m_B(P \cup F)$	ϕ	F	$P \cup F$
$m_B(\Theta)$	D	$F \cup M$	Θ

and a total mass of 0.15 *is lost on the empty set. This means that the assignments should be rescaled by a factor* $K = 1/0.85 = 1.1765$ *(to four decimal places). The mass matrix becomes,*

	$m_A(D)$	$m_A(F \cup M)$	$m_A(\Theta)$
$m_B(D \cup M)$	0.0882	0.1176	0.0882
$m_B(P \cup F)$	0.1765	0.2353	0.1764
$m_B(\Theta)$	0.0882	0.1176	0.0882

and the calculation for the belief functions and uncertainty intervals proceeds exactly as before.

The differences between the DS approach and the probabilistic are now manifest. First of all, probabilistic—or rather Bayesian—approaches are unable to accommodate uncertainty. All probability must be assigned to the set of propositions under consideration. Secondly, the Bayesian approach is unable to meaningfully assign probabilities to the union of propositions. If the uncertainty for all propositions is zero and the mass assigned to unions is zero, DS is reduced to Bayesian probability reasoning.

There are other frameworks that seek to extend Bayesian methods in a similar manner to DS such as the generalized evidence processing (GEP) approach of [5] and those proposed in [50, 51].

5.2 The DS neural network

The object of this section is to briefly describe the neural network implementation of the DS-based

classifier. More detail can be found in the original reference [46].

The basic idea is to assign one of M classes $C_1, \ldots, C_M$ (these form the frame of discernment) to a feature vector $\underline{x}$ on the basis of a set of N training examples $\underline{x}_1, \ldots, \underline{x}_N$. Suppose the vector $\underline{x}$ is close to a training example $\underline{x}_i$ with respect to an appropriate distance measure d ($d_i = \|\underline{x} - \underline{x}_i\|$). It is then appropriate that the class of the vector $\underline{x}_i$ influences ones beliefs about the class of $\underline{x}$. One has evidence about the class of $\underline{x}_i$. The approach to the classification taken in [48] is to allocate belief to the event C_q (the class of $\underline{x}_i$, according to the distance d_i).

$$m^i(C_q) = \alpha\phi_q(d_i) \tag{27}$$

where $0 < \alpha < 1$ is a constant and ϕ_q is an appropriate monotonically decreasing function. Each training vector close to $\underline{x}$ contributes some degree of belief. For each training vector, a degree of belief is also assigned to the whole frame of discernment Θ as follows:

$$m^i(\Theta) = 1 - \alpha\phi_q(d_i) \tag{28}$$

The function ϕ used here is the basic Gaussian,

$$\phi_q(d_i) = \exp(-\gamma_q(d_i)^2) \tag{29}$$

where γ_q is a positive constant associated with class q. To simplify matters, one confines the construction of the belief assignment for the vector $\underline{x}$ to a sum of the beliefs induced by its nearest neighbors. The sum is computed using Dempster's combination rule as described in the previous subsection. Actually, a further simplification is made to speed up the processing. Rather than summing over the nearest neighbors from the whole training set in order to assign the belief, one sums over a set of *prototypes* constructed from the training set by a clustering algorithm. Each prototype $\underline{p}_i$ is assigned a degree of membership to the class q denoted by u_q^i with the constraint $\sum_{q=1}^{M} u_q^i = 1$. These are used to compute the belief in the class q for $\underline{x}$ given the distances d_i from the prototypes

Although it is a gross simplification, the algorithm can be summed up as follows:

1. Construct the prototypes $\underline{p}_i$ from the training data using a clustering algorithm.
2. Given a vector $\underline{x}$, compute the distances from the vector to the prototypes. Using the parameters d_i and u_q^i, assign a degree of belief for each class q.
3. Use Dempster's combination rule to compute the total belief in each class from all the contributing prototypes.

The algorithm extends the probabilistic classifier by also making an assignment to the frame of discernment and this quantifies the degree of uncertainty of the classification. Reference [48] explains how the algorithm can be implemented in terms of a four-layer neural network. Unlike an MLP, the network is not a simple feed-forward structure.

In order to assign a class to the vector $\underline{x}$, one selects that with the largest overall belief assignment induced by the training data.

5.3 Damage location example

The same problem that was investigated in the first case study, namely damage location on a Gnat aircraft wing (Figure 2), is revisited in order to compare the evidence-based approach to damage classification with the probabilistic approach. The only significant difference between the data acquisition and signal processing methodology is that outlined in the previous •chapter and that used in the current study related to the number of features used. In the previous study, the candidate features were reduced to form a feature set comprising of nine features. In the current study, the candidate features were reduced to form a feature set of four individual features. The best four features were selected from the initial set using a GA to optimize the classification error of an MLP classifier [52]. The reduction to four features was made to ensure that the MLP network used for comparison with the DS network later was unlikely to suffer from overtraining. Q8

By performing the novelty analysis for each of the four features, a data matrix of size [1800 × 4] was obtained. As before, this was divided into three equal parts to form separate *training, validation*, and *test* data sets for subsequent network evaluation. Finally, the data sets were logarithmically compressed and normalized to −1 to +1 before presentation

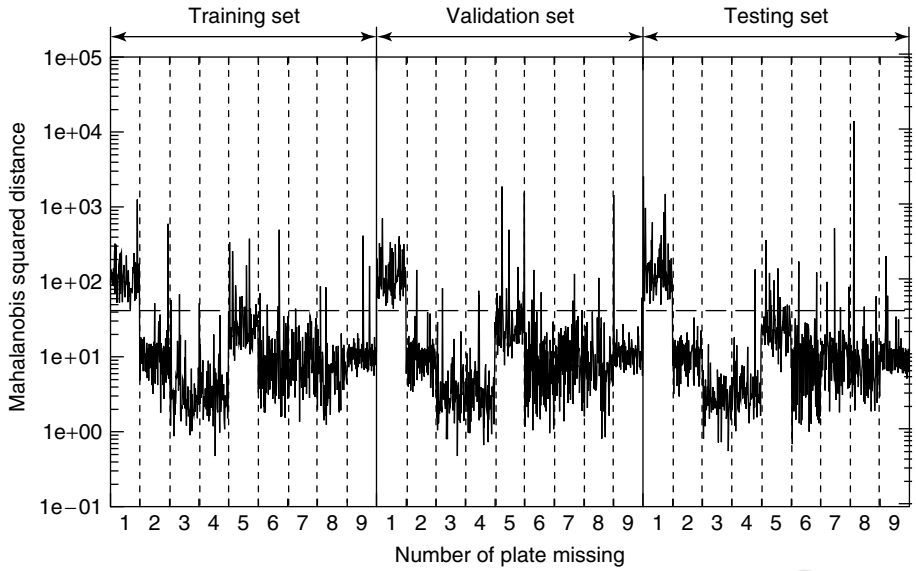

Figure 8. Outlier statistic for all damage states for the novelty detector trained to recognize panel 1 removal.

to the network. Figure 8 shows the results of the outlier analysis for a feature that was clearly able to recognize the removal of inspection panel 1.

The plot in Figure 8 shows the discordancy (novelty index) values returned by the novelty detector over the whole set of damage states. The horizontal dashed lines in the figures are the thresholds for 99% confidence in identifying an outlier; they are calculated according to the Monte Carlo scheme described in [53]. The novelty detector substantially fires only for the removal of panel for which it has been trained. The outputs from the four novelty detectors are then used to form the data with which the damage location classifiers are trained.

5.4 DS network results

The final stage of the analysis was to produce a classifier based on the DS neural network algorithm, which could serve as a damage location system. As with a standard MLP network, the specification of the DS network structure requires hyperparameters; in this case, the number of prototypes (analogous to the number of hidden units in the first layer of the network) and the starting values of the weights before training. These were computed by a cross-validation procedure as for the MLP [54]. Many neural networks were trained with the same training data but with differing numbers of prototypes and initial weights. Up to 30 prototypes were considered, and in each case 10 randomly chosen initial conditions were used. The best network was selected by observing which of them produced the minimum misclassification error on the validation set. The final judgment of the network capability was made by using the independent testing set.

The results for the presentation to the classifier are summarized in the confusion matrix given in Table 1. The best DS network used 29 prototypes. The probability of correct classification was 89.7%. There were four events associated with the frame of discernment, corresponding to a probability mass of 0.007. This means that, allowing for the fact that the network indicates when it has insufficient evidence to make a classification, the classification error is 9.6%. The main source of confusion is in locating damage to the two smallest panels, 3 and 6, which was anticipated.

In order to make a comparison with the standard approach, the algorithm chosen was a standard MLP neural network. The neural network was presented with the same four novelty indices at the input layer and required to predict the damage class at the output layer. The procedure for training the neural

Table 1. Confusion matrix for the best DS network using four log features–testing set

Prediction	1	2	3	4	5	6	7	8	9	Θ
True class 1	54	5	5	0	0	0	2	0	0	0
True class 2	0	63	0	0	2	0	0	0	0	1
True class 3	6	1	56	2	0	0	0	0	0	1
True class 4	5	0	1	55	0	3	0	2	0	0
True class 5	0	0	0	0	65	0	0	1	0	0
True class 6	2	2	2	4	0	54	1	0	0	1
True class 7	0	1	1	0	0	0	61	2	1	0
True class 8	0	0	1	0	1	0	0	62	1	1
True class 9	0	0	0	0	0	0	0	3	63	0

network again followed the guidelines in [54]. The training set was used to establish weights, whilst the network structure and training time etc. were optimized using the validation set. The testing set was then presented to this optimized network to arrive at a final classification error. For the network structure, the input layer necessarily had four neurons, one for each novelty index, and the output layer had nine nodes, one for each class. The training phase used the same 1 of M strategy as in the previous case study.

The best neural network had 19 hidden units and resulted in a testing classification error of 0.118 i.e., 89.2% of the patterns were classified correctly. This means that the misclassification probability is 10.8%. The confusion matrix is given in Table 2. Again, the main errors were associated with the two small panels P3 and P6.

The main conclusion here is that the DS approach to classification implemented as a neural network gives comparable results to the standard analysis using an MLP neural network. The data used is from a full-scale experimental test on an aircraft wing and is therefore a stringent test of algorithms from the point of view of SHM. In fact, the DS network shows a slight improvement, giving a classification probability of 89.7% compared to 89.2% for the MLP. One possible advantage of the DS approach is the fact that it can assign patterns to the frame of discernment and thus indicate to the analyst that there is insufficient evidence to make a classification. Another advantage of using the DS approach might be to provide methodological diversity in order to use data fusion. In this case, the errors made by the DS network and the MLP network are slightly different and therefore there is some advantage to fusion—this is considered in [55].

6 CONCLUSIONS

The purpose of this article was to introduce the many frameworks for the representation of uncertainty and the three uncertainty-related issues of interest in the field of structural dynamics. In order to put these ideas into context, and to showcase the merits of nonprobabilistic treatments of uncertainty, within the area of SHM, two SHM-related case studies were discussed.

Table 2. Confusion matrix for best MLP neural network using four log features–testing set

Prediction	1	2	3	4	5	6	7	8	9
True class 1	61	1	0	0	0	0	1	0	0
True class 2	0	63	0	0	3	0	0	0	0
True class 3	1	0	48	8	0	5	7	2	0
True class 4	0	1	3	56	0	2	0	4	0
True class 5	0	0	0	0	66	0	0	0	0
True class 6	4	1	0	9	0	52	0	0	0
True class 7	1	0	0	0	0	0	59	5	1
True class 8	1	0	0	0	1	0	1	59	4
True class 9	0	0	0	0	0	0	0	5	61

The first case study illustrated how, by taking an alternative approach to the convention technique of neural network selection via cross-validation, the robustness of a network to data uncertainty could be drastically improved. This finding should have implications related to the issue of certification of neural network classifiers and regressors for the purpose of monitoring safety-critical structures. The second case study illustrated that the use of an evidence-based classifier, with its ability to admit ignorance, could reduce the likelihood of misclassifications within an SHM system. This will clearly have both economic and safety-related benefits.

REFERENCES

[1] Hemez FM, Doebling •SW. *Uncertainty, Validation of Computer Models and the Myth of Numerical Predictability*, LANL Report LA-UR-01-2492, 2001.

[2] Rasmussen NC, *et al. Reactor Safety Study: An Assessment Accident Risks in U.S. Commercial Nuclear Power Plants*. US Nuclear Regulatory Commission, NUREG-75/014, WASH-1400, 1975.

[3] Dempster AP. Upper and lower probabilities induced by a multi-valued mapping. *Annals of Mathematical Statistics* 1967 **38**:325–339.

[4] Shafer G. *A Mathematical Theory of Evidence*. Princeton University Press, 1976.

[5] Thomopoulos SCA. Theories in distributed decision fusion: comparison and generalisation. *Sensor Fusion III: 3D Perception and Recognition. Proceedings of SPIE*.1383 1990.

[6] Oberkampf WL, Helton JC, Sentz K. Mathematical representation of uncertainty. *AIAA Non-Deterministic Approaches Forum*, AIAA 2001-1645. Seattle, WA, 2001.

[7] Osegueda RA, Lopez H, Pereyra L, Ferregut CM. Localisation of damage using fusion of modal strain energy differences. *Proceedings of 18th International Modal Analysis Conference*, IMAC XVIII. San Antonio, TX, 2000; pp. 695–701.

[8] Pereyra L, Osegueda RA, Carrasco C, Ferregut CM. Detection of damage in a stiffened plate from fusion of modal strain energy differences. *Proceedings of 18th International Modal Analysis Conference*, IMAC XVIII. San Antonio, TX, 2000; pp. 1556–1562.

[9] Dubois D, Prade H. *Possibility Theory: An Approach to Computerised Processing of Uncertainty*. Plenum Press: New York, 1986.

[10] Klir GJ. On fuzzy-set interpretation of possibility theory. *Fuzzy Sets and Systems* 1999 **108**:263–273.

[11] Zadeh LA. Fuzzy sets. *Information and Control* 1965 **8**:338–353.

[12] Lallemand B, Plessis G, Tison T, Level P. Modal behaviour of structures defined by imprecise geometric parameters. *Proceedings of 18th International Modal Analysis Conference*, IMAC XVIII. San Antonio, TX, 2000; pp. 1422–1428.

[13] Plessis G, Lallemand B, Tison T, Level P. A fuzzy method for the modal identification of uncertain experimental data. *Proceedings of 18th International Modal Analysis Conference*, IMAC XVIII. San Antonio, TX, 2000; pp. 1831–1837.

[14] Plessis G, Lallemand B, Tison T, Level P. Fuzzy modal parameters. *Journal of Sound and Vibration* 2000 **233**:797–812.

[15] Lallemand B, Cherki A, Tison T, Level P. Fuzzy modal finite element analysis of structures with imprecise material properties. *Journal of Sound and Vibration* •2000 353–364.

[16] Moore RE. *Methods and Applications of Interval Analysis*. SIAM: Philadelphia, PA, 1979.

[17] Muhanna RL, Mullen RL. Formulation of fuzzy finite element methods for mechanics problems. *Computer-Aided Civil and Infrastructure Engineering* 1999 **14**:107–117.

[18] Worden K, Osegueda R, Ferregut C, Nazarian S, George DL, George MJ, Kreinovich V, Kosheleva O, Cabrera S. Interval methods in non-destructive testing of material structures. *Reliable Computing* 2001 **7**:341–352.

[19] Comba JLD, Stolfi J. •Affine arithmetic and its applications to computer graphics. *Proceedings of SIBGRAPH '93*, 1993; pp. 9–18.

[20] Manson G. Sharper eigenproblem estimates for uncertain multi degree of freedom systems. *Proceedings of 21st International Modal Analysis Conference*. Orlando, FL, 2003.

[21] Ben-Haim Y, Elishakoff I. *Convex Models of Uncertainty in Applied Mechanics*. Elsevier Science: 1990.

[22] Ben-Haim Y. Set-models of information-gap uncertainty: axioms and an inference scheme. *Journal of the Franklyn Institute* 1999 **336**:1093–1117.

[23] Hemez FM, Ben-Haim Y, Cogan S. *Information-Gap Robustness for the Test-Analysis Correlation of a Non-Linear Transient Simulator*. LA-UR-02-3538. Los Alamos National Laboratory Report, 2002.

Q12

[24] Klir •GJ, Smith RM. On measuring uncertainty and uncertainty-based information. *Annals of Mathematics* 2001.

[25] Choquet G. Theory of capacities. *Annales de L'Institut Fourier* 1953-54 **5**:132–295.

[26] Reese CS, Wilson AG, Hamada MS, Martz HF, Ryan KJ. *Integrated Analysis of Computer and Physical Experiments*. LA-UR-00-2915. Los Alamos National Laboratory Report, 2000.

[27] Bement TR, Booker JM, Sellers KF, Singpurwalla ND. *Membership Functions and Probability Measures of Fuzzy Sets*. LA-UR-00-3660. Los Alamos National Laboratory Report, 2000.

[28] Bergman LA. Uncertainty modelling in dynamical systems: a perspective. In *Structural Dynamics @ 2000: Current Status and Future Directions*, Ewins DJ and Inman DJ (eds). Research Studies Press: Baldock, England, 2001, pp. 9–16.

[29] Manson G, Worden K, Allman DJ. Experimental validation of a structural health monitoring methodology. Part II. Novelty detection on a Gnat aircraft. *Journal of Sound and Vibration* 2003 **259**(2):345–363.

[30] Manson G, Worden K, Allman DJ. Experimental validation of a structural health monitoring methodology. Part III. Damage location on an aircraft wing. *Journal of Sound and Vibration* 2003 **259**(2):365–385.

[31] Bishop CM. *Neural Networks for Pattern Recognition*. Oxford University Press, 1995.

[32] Nabney IT. *Netlab-Algorithms for Pattern Recognition*. Springer-Verlag, 2002.

[33] Haykin S. *Neural Networks, a Comprehensive Foundation*, *Second Edition*, Prentice Hall, 1999.

[34] MacKay DJC. *Information Theory, Inference, and Learning Algorithms*. Cambridge University Press, 2003.

[35] MacKay DJC. The evidence framework applied to classification networks. *Neural Computation* 1992 **4**(5):720–736.

[36] Papadopoulos G, Edwards PJ. Confidence estimation methods for neural networks: a practical comparison. *IEEE Transactions on Neural Networks* 2001 **12**(6):1278–1287.

[37] Lowe D, Zapart C. Point-wise confidence interval estimation by neural networks: a comparative study based on automotive engine calibration. *Neural Computing and Applications* 1999 **8**:77–85.

[38] Ben-Haim Y, Elishakoff I. *Convex Models of Uncertainty in Applied Mechanics*. Elsevier, 1990.

[39] Ben-Haim Y. *Robust Reliability in the Mechanical Sciences*. Springer-Verlag, 1996.

[40] Hemez FM, Ben-Haim Y, Cogan S. Information gap robustness for the test-analysis correlation of a nonlinear transient simulation. *Proceedings of the 9th AIAA/ISSMO Symposium on Multi-disciplinary Analysis and Optimisation*. Atlanta, GA, 2002.

[41] Ben-Haim Y. *Information-Gap Decision Theory*. Academic Press, 2001.

[42] Moore RM. *Interval Analysis*. Prentice Hall, 1966.

[43] Worden K, Tomlinson GR. *Nonlinearity in Structural Dynamics*. Institute of Physics Publishing, 2001.

[44] Rump SM. INTLAB – INTerval LABoratory. In *Developments in Reliable Computing*, Csendes T (ed). Kluwer Academic Publishers, 1999, pp. 77–104.

[45] Pierce SG, Worden K, Manson G. Information-Gap analysis of a neural network damage locator, *IMechE meeting on Pattern Recognition-Detection, Classification and Monitoring*. University of Liverpool: Liverpool, 2004.

[46] Hayton P, Utete S, King D, King S, Anuzis P, Tarassenko L. Static and dynamic novelty detection methods for jet engine health monitoring. *Philosophical Transactions of the Royal Society A: Mathematical Physical and Engineering Sciences* 2007 **365**:493–514.

[47] Worden K, Manson G. The application of machine learning to structural health monitoring. *Philosophical Transactions of the Royal Society A: Mathematical Physical and Engineering Sciences* 2007 **365**:515–537.

[48] Denoeux T. A neural network classifier based on Dempster-Shafer theory. *IEEE Transactions on Systems, Man, and Cybernetics* 2000 **30**:131–150.

[49] Shafer G, Logan R. Implementing Dempster's rule for hierarchical evidence. *Artificial Intelligence* 1987 **33**:271–298.

[50] Barnett JA. Computational methods for a mathematical theory of evidence. *Proceedings of the 7th International Joint Conference on Artificial Intelligence*, 1981; pp. 868–875.

[51] Yen J. A reasoning model based on an extended Dempster-Shafer theory. *Proceedings of the 5th AAAI-86 National Conference on Artificial Intelligence*, 1986; pp. 125–131.

[52] Worden K, Manson G, Hilson G, Pierce SG. Genetic optimisation of a neural damage locator. *Journal of Sound and Vibration* 2007.

[53] Worden K, Manson G, Fieller NR. Damage detection using outlier analysis. *Journal of Sound and Vibration* 2000 **229**:647–667.

[54] Tarassenko L. *A Guide to Neural Computing Applications*. Arnold, 1998.

[55] Worden K, Manson G, Denoeux T. *An Evidence-Based Approach to Damage Localisation on an Aircraft Structure*. •Submitted to Mechanical Systems and Signal Processing 2007. Q13

FURTHER READING

Goldberg DE. *Genetic Algorithms in Search, Optimisation and Machine Learning*. Addison-Wesley, 1989.

Please note that the abstract and keywords will not be included in the printed book, but are required for the online presentation of this book which will be published on Wiley InterScience (www.interscience.wiley.com). If the abstract and keywords are not present below, please take this opportunity to add them now.

The abstract should be of short paragraph upto 200 words in length and keywords between 5 to 10 words.

Abstract: This article considers the issue of uncertainty analysis in general and also its relevance to structural health monitoring. Brief descriptions of the most popular of the many frameworks for uncertainty representation are given. The three main uncertainty-related problems of relevance to structural dynamics, namely, quantification, fusion, and propagation, are then discussed. In order to illustrate the preceding ideas in a realistic scenario, two case studies conducted on an aerospace structure, namely the wing of a Gnat trainer aircraft, are provided. The first case study considers the issue of attaining certification for the artificial neural network damage classifiers through the assessment of the network's robustness to uncertainty. This case study involves the propagation of intervals through the network structure and it was found that it is likely that the networks which would be considered as optimal, in the traditional sense, would not be the network that is most robust to uncertainty. The second case study considers evidence-based classifiers as an alternative to probabilistic classifiers for the problem of damage location. The Dempster–Shafer theory is employed to construct neural network classifiers with the potential to admit ignorance, rather than misclassify. Issues of propagation and fusion in an evidence-based framework are considered. It was also found that Dempster–Shafer networks give a slight improvement over their probabilistic counterparts.

Keywords: uncertainty; probability theory; evidence theory; fuzzy logic; possibility theory; interval; convex models; propagation; quantification; fusion; damage location.

QUERIES TO BE ANSWERED BY AUTHOR (SEE MARGINAL MARKS Q..)

IMPORTANT NOTE: You may answer these queries by email. If you prefer, you may print out the PDF, and mark your corrections and answers directly on the proof at the relevant place. Do NOT mark your corrections on this query sheet. Please see the proofing instructions for information about how to return your corrections and query answers.

Q1. Please spell out the initials for this author.
Q2. Please confirm if the labels in Figure 4 are fine.
Q3. Please confirm twocolumn table (page no 16 bottom) for pagination purpose.
Q4. This article does not contain any cross-references to other articles in the Encyclopedia. The editors and publishers believe that cross-references are very important for this publication. Please consider whether you could add any cross-references to other chapters. Details of how to provide cross-references are included in your proofreading instructions.
Q5. Please confirm if this abbreviation "SEA" needs to be spelt out. If yes, please provide the expansion.
Q6. Please confirm if this abbreviation "DSTL" needs to be spelt out. If yes, please provide the expansion.
Q7. In the sentence 'On the basis of sensor evidence, each event or union of events is assigned a degree of probability mass or . . .' please resupply the variable preceding 'is the empty set' as there is a font missing.
Q8. In the sentence 'The only significant difference between the data acquisition and signal processing methodology is that outlined in the previous chapter and . . .' please provide the article ID for the article (chapter) referred to here.
Q9. Please provide the publisher's name for this reference.
Q10. Please provide the volume number for reference 15.
Q11. Please provide the place of conference for references 19, 50 and 51.
Q12. Please provide the volume number and page range for references 24 and 52.
Q13. Please clarify if this article has since been published. If so, please provide the complete details for reference 55.