

Statistical and Dempster-Shafer techniques in testing structural integrity of aerospace structures

Roberto A. Osegueda, Seetharami R. Seelam, Bharat Mulupuru, and Vladik Kreinovich
FAST Center, University of Texas, El Paso, TX 79968, USA

ABSTRACT

Several techniques are known for non-destructive testing of aerospace structures, such as pulse echo, Eddy current, magnetic resonance, etc. Each of these techniques detects some faults but misses others, so it is desirable to combine (fuse) the results of these techniques. Several methods of data fusion are known. To improve the quality of fault detection, we modified the straightforward statistical method as follows: (1) we computed mean and variance iteratively: detected faults are excluded from the computation on the next iteration; (2) we treated the plate's edge and the inside separately; (3) we dismissed measurements in which only one technique detects a fault as possibly erroneous. The resulting method indeed leads to a much better fault detection.

Keywords: fault detection, statistical methods, Dempster-Shafer methods

1. INTRODUCTION

Testing structural integrity of aerospace structures is an important practical problem (see, e.g., Ferregut et al.²). Some of the faults can be detected visually, but other (sometimes potentially dangerous) faults do not show up during the visual inspection. To detect such faults, we must use some indirect (non-destructive) testing techniques.

Several such techniques are known, such as pulse echo, Eddy current, magnetic resonance, etc. All these techniques aim at determining whether there is a fault, and if there are faults, what is the location and the size of each fault. Each of these techniques detects some faults but misses others, so it is desirable to combine (fuse) the results of these techniques. In addition to detecting faults, all the images produced by these techniques come with a lot of noise; therefore, the desired fusion technique must also decrease the noise level.

2. TEST STUDY

To test ability of different NDE techniques, Boeing developed a benchmark plate. It is a 11" \times 11" B-52 plate which contains artificially induced smooth-shaped (circular) and angular-shaped (square) faults of four different sizes, both inside and on the edge. In total, we have 16 faults: 8 squares and 8 circles. Of these 16 faults, 8 are inside the plate: four squares of sides 1/2", 3/8", 1/4", and 1/8"; and four circles of diameters 1/2", 3/8", 1/4", and 1/8"; and 8 are of the exact same size but on the edge of the plate: four squares of sides 1/2", 3/8", 1/4", and 1/8"; and four circles of diameters 1/2", 3/8", 1/4", and 1/8".

On this plate, the following quantities are measured: Pulse Echo (PE0 and PE1) corresponding to different time delays, Eddy Current (EC), and Magnetic Resonance (RES0, RES1, RES2, and RES3). PE and RES return two components – real and imaginary – for each point. Empirical analysis shows that considering the magnitudes of the corresponding complex values does not improve the quality of fault detection. Therefore, for each of these complex-valued measurements, we only consider phases.

Further author information: (Send correspondence to V.K.)
V.K.: E-mail: vladik@cs.utep.edu, Telephone: 1 915 747 6951

3. EXISTING METHODS OF FAULT DETECTION

3.1. Fault Detection Based on a Single Image (i.e., on Measurement of a Single Type)

The general idea of the traditional approach to fault detection is as follows. For each image i , we compute the (arithmetic) *average* a_i of all the values, and the mean square (standard) deviation σ_i of the signal $x_i(A)$ from its average a_i . In accordance with the standard statistical techniques (see, e.g., Wadsworth¹⁵), we can then consider points A at which the value $x_i(A)$ is outside of the “two sigma” interval $[a_i - 2\sigma_i, a_i + 2\sigma_i]$ as possible faults. In other words, we compute the “normalized” values $z_i(A) = (x_i(A) - a_i)/\sigma_i$, and consider points A for which $|z_i(A)| > 2$ to be possible fault locations.

By applying this idea to each of the seven image, we had the following results:

- PE0 well detected edge faults, and not other faults.
- PE1 detected the boundaries of square edge faults and all circular faults.
- EC0 detected inside circular faults and nothing else.
- The phase of RES0 detected all circular faults near the plate’s edge.
- The phases RES1 and RES2 also detected (but more fuzzily) all circular faults near the plate’s edge.
- The phase of RES3 detected all square faults.

In addition to detecting the faults, all these methods lead to several “false positives”, i.e., no-fault points erroneously marked as faults. In particular, practically all of these methods marked several points on the edge of the plate as faults.

3.2. The Need for Data Fusion

In short, none of the seven images detected all the faults. However, every fault was detected by at least one image. It is therefore desirable to combine (fuse) the information contained in these seven images. We tried all fusion methods described in Gros.³

3.3. Union Method

The first – *union* – method is motivated by the fact that since every fault is detected by one of the measurements, it is natural to declare a point a fault if at least one of the 7 measurements indicates that this point is a fault. For this method, 195 out of 665 fault points were undetected (*false negatives*); also, we have 3,470 false positives.

3.4. Average Method

In the union method, to decide whether a give point A is a fault, we only used one bit of information about each of 7 measurements i : whether this point A is classified as a fault by this measurement, i.e., whether $|z_i(A)| > 2$. To improve the quality of data fusion, it is therefore desirable to use one of the *continuous* fusion methods that combines the actual values $z_i(A)$ into a single “fused” value $z(A) = f(z_1(A), \dots, z_n(A))$, and then compares the resulting value $z(A)$ with a threshold. If one of the values $z_i(A)$ is very large, the point is probably a fault; problems appear when the values $z_i(A)$ are reasonably small; for such points, we can ignore quadratic and higher order terms in the expansion of f and consider f to be linear: $f = \sum w_i \cdot z_i$ for some weights w_i . Since there is no reason to believe that one of the 7 measurements is better than the others, it makes sense to consider equal weights w_i . In this case, the use of the fused value is equivalent to using the (arithmetic) *average* $\bar{z}(A)$ of $z_i(A)$. The standard deviation of an average of n values is $1/\sqrt{n}$, so, following 2σ rule, we mark A as a fault if $|\bar{z}(A)| > 2/\sqrt{n}$.

This *average method* leads to 510 false negatives and 778 false positives: fewer false positives but more false negatives than for the union. To compare the union and average methods, we can use $k \cdot \sigma$ for $k \neq 2$: e.g., fix the number of false negatives, for each method, choose $k \neq 2$ that leads to exactly this many false negatives FN, and compare the resulting number of false positives. Depending on FN, the average method produces slightly better or slightly worse results, but never much better ones.

3.5. Probabilistic, Neural, and Fuzzy Approaches

Gros³ describes three other approaches to information fusion: probabilistic, neural, and fuzzy. To apply *probabilistic* or *neural* methods to detect faults on the plate, we need to collect a sufficient amount of data about the faults. In aging and futuristic aircraft design we do not have this information:

- For aging aircraft, B-52 was designed 50 years ago. We have statistics from the aircraft for the last 50 years. After these 50 years of service, to detect faults on the aircraft for the next 3 years, we cannot use the statistics we have for the last 50 years, because during the service the aircrafts material properties alter because of the wear, tear, and cracks.
- In case of futuristic aircraft design, these designs are entirely different from the existing designs, so we cannot use the statistics we have from the other aircrafts to detect faults on our new ones.

Fuzzy logic converts expert knowledge (usually, rules), expressed in natural language, into computer algorithms. For fault detection in aging aircraft, experts cannot provide us with meaningful rules, so we cannot use this approach.

3.6. Dempster-Shafer Approach

Another approach described in Gros³ is Dempster-Shafer (DS) approach. In this method, in the general spirit of statistics, for every measurement i and for every point A , we use the value $z_i(A)$ not just to mark or not to mark A as a fault, but rather to gauge the probability $p_i(A)$ that A is a fault. The larger $z_i(A)$, the more probable it is that A is a fault, so $p_i(A) = f_i(z_i(A))$ for some increasing function $f(z)$. The simplest function is a linear function. When $x_i(A) = a_i$ and $z_i(A) = 0$, it is most certainty not a fault; when $z_i(A)$ takes its largest possible value Z_i , it is most probably a fault. So, $f_i(z)$ is a linear function with $f(0) = 0$ and $f(Z_i) = 1$, hence $f(z) = z/Z_i$.

In many practical measurements, deviations from average are only in one direction of increasing $x_i(A)$ and hence, $z_i(A)$. For such measurements, we can safely assume that values $x_i(A) < a_i$ do not represent fault, i.e., that $f_i(z) = 0$ for all $z < 0$.

We can now compute the plausibility $Pl(A)$ that point A is an outlier point as the probability that at least one of the measurements detected point A as an outlier:

- the probability that, according to i -th measurement, point A is not an outlier point, is equal to $1 - p_i(A)$;
- since different measurements detect different types of faults, it is natural to assume that the measurements are statistically independent; then, the probability that point A *is not* an outlier point according to all the measurements is equal to the product $\prod_i (1 - p_i(A))$;
- finally, the probability that point A is an outlier point according to at least one measurement is equal to $Pl(A) = 1 - \prod_i (1 - p_i(A))$.

We can also compute the *belief* $Bel(A)$ that there is a fault at a point A as the probability that all the measurements detected the fault. Since the measurements are independent, this probability is equal to the product $Bel(A) = \prod_i p_i(A)$. The actual probability of A being a fault is somewhere within the interval $[Bel(A), Pl(A)]$. Gros³ recommends to use the midpoint of this interval as the reasonable estimate for the probability. Alas, the use of this midpoint leads to a much worse fault detection than union or average (even if we use the actual CDF instead of a linear function $f(z)$).

3.7. Conclusion

So, among the existing methods, union and average are the best. Problem: there is still too many false positives and false negatives, so we need a better fault detection method.

4. A NEW VERSION OF A DEMPSTER-SHAFER APPROACH

While analyzing why the traditional DS approach did not work, we noticed (see, e.g., Holguin⁴) that the results are much better if as an estimate for the probability of a fault $p(A)$, instead of the midpoint, we take the upper (worst-case) endpoint $Pl(A)$ of this interval. In other words, we consider a point A a fault if $Pl(A)$ exceeds a certain threshold t .

To compare this method with union and average methods, we fix the number of FN, change k (or threshold t) in each method to get exactly this FN, and compare the number of false positives. By this comparison, the DS method is better – but only slightly better – than union and average: e.g., for FN=75, we get 7022 FP for union, 6941 for average, and only 6806 for DS. How can we further improve fault detection?

5. OUR FIRST IDEA: ITERATIVE APPROACH

The main drawbacks of the existing methods is that some faults are undetected. One reason is that we use 2σ criteria, and in the computation of the average a_i and of the standard deviation σ_i we include all the points A including fault points, in which the value of the signal is very different from the expected normal value.

Hence, the computed σ_i is much larger than for the normal points. As a result, some true fault points – which would have been detected by a correct (smaller) σ_i go undetected when we use this larger value.

In our test case, for some measurements, the values of $x_i(A)$ corresponding to most actual fault locations are within $2\sigma_i$ from the average. Thus, the resulting points are not recognized as fault locations.

To detect all the faults, we apply the iterative estimation:

- first, we use the traditional procedure and find some possible faults;
- then, we re-calculate a_i and σ_i by omitting points known to be possible faults;
- we then check for new faults by using these better estimates for a_i and σ_i ,

and we continue these iterations until the subsequent values of $a_i^{(s)}$ and $\sigma_i^{(s)}$ become close to each other, e.g., when $|a_i^{(s-1)} - a_i^{(s)}| \leq 0.1 \cdot a_i^{(s)}$ and $|\sigma_i^{(s-1)} - \sigma_i^{(s)}| \leq 0.1 \cdot \sigma_i^{(s-1)}$.

As a result, we arrive at the following iterative algorithm. At each step $s = 1, 2, \dots$ of this algorithm:

- new approximations $a_i^{(s)}$ and $\sigma_i^{(s)}$ are computed for the desired values of a_i and σ_i ; and
- some points on the plate are marked as possible fault locations.

In the beginning, we do not know where the faults are, so no points are marked. Each step s consists of the following two operations:

- first, we compute the average $a_i^{(s)}$ and the standard deviation $\sigma_i^{(s)}$ based on all un-marked points of the plate;
- then, we check each un-marked point, and mark all the points for which $|x_i(A) - a_i^{(s)}| > 2\sigma_i^{(s)}$.

The process stops when the subsequent values of $a_i^{(s)}$ and $\sigma_i^{(s)}$ become close to each other, i.e., when

$$|a_i^{(s-1)} - a_i^{(s)}| \leq 0.1 \cdot a_i^{(s)} \quad \text{and} \quad |\sigma_i^{(s-1)} - \sigma_i^{(s)}| \leq 0.1 \cdot \sigma_i^{(s-1)}.$$

This process is guaranteed to stop for the following reason: Once marked, a point remains marked; so the set of all un-marked points can only shrink. We start with a finite number of un-marked points, so this shrinking process has to stop.

The corresponding final values of $a_i^{(s)}$ and $\sigma_i^{(s)}$ are then taken as the desired values of a_i and σ_i . As a result, for each measurement i and for each point A , we get a normalized value $z_i(A) = (x_i(A) - a_i)/\sigma_i$.

Our histogram analysis confirmed that the value which are ≤ 2 (i.e., which correspond to probably no-fault locations) are distributed according to Gaussian distribution with the mean 0 and standard deviation 1. Thus, for these points, the use of “two sigma” criterion is indeed justified.

We can use this ideas in all three methods: union, average, and DS. In all three cases, we get better fault detection, but still not perfect one: all edge points are marked as faults. So, we need to continue improving the method.

6. INFORMATION FUSION: TRADITIONAL STATISTICAL APPROACH

We have mentioned that we could not apply probabilistic data fusion methods because we did not know the probabilities of different faults. However, now, since we know the distribution (Gaussian) of $z_i(A)$ for non-fault points and we have assumed that the distributions corresponding to different measurements are independent, we can use statistical data fusion.

For each measurement i , the distribution for non-fault points is Gaussian with 0 mean and unit standard deviation. According to this normalized Gaussian distribution, for each value z_i , the probability density is proportional to $\exp(-(z_i(A))^2/2)$. Due to independence, the probability (density) that a non-fault point A can have the normalized measured values $z_1(A), \dots, z_n(A)$ is equal to the product of the probabilities corresponding to different measurements, and is, therefore, proportional to

$$\prod_{i=1}^n \exp(-(z_i(A))^2/2) = \exp\left(-(1/2) \cdot \sum_{i=1}^n (z_i(A))^2\right).$$

If this probability is too small, this means that we probably have a fault. Thus, the criterion for a fault is that the above probability is $\leq \text{const}_1$ for some constant const_1 . By turning to logarithms, we can reformulate this criterion is the equivalent simpler form $\sum (z_i(A))^2 \geq \text{const}_2$, or, equivalently, $(1/n) \cdot \sum_{i=1}^n (z_i(A))^2 \geq \text{const}_3$. To complete our description, it is therefore sufficient to select an appropriate constant const_3 .

In accordance with the traditional statistics, we should choose the threshold value const_3 in such a way that for no-fault points, the probability of exceeding const_3 is $\approx 5\%$ (this corresponds to “two sigma” criterion).

For large n , according to the central limit theorem, the average $\overline{z^2(A)}$ of the values $(z_i(A))^2$ is distributed approximately according to the normal distribution. The mean $E[\overline{z^2}]$ of this distribution is equal to the mathematical expectation $E[z_i^2]$ of z_i^2 , i.e., to $E[z^2] = \sigma^2[z] = 1$. The mean square $E[(\overline{z^2} - E(\overline{z^2}))^2] = \sigma^2[\overline{z^2}]$ of the difference between the average $\overline{z^2}$ and its mean 1 is equal to $(1/n)$ -th of the mean square

$$E[(z^2 - 1)^2] = E[z^4] - 2 \cdot E[z^2] + 1 = 3\sigma^4 - 1 = 2$$

(see, e.g., Wadsworth¹⁵). Thus, the average $\overline{z^2}$ has the mean 1 and the standard deviation $\sqrt{2/n}$. Hence, if $\overline{z^2(A)} \geq 1 + 2 \cdot \sqrt{2/n}$, then we can be sure that there is fault at this point.

So, the criterion for declaring a point a fault location is $(1/n) \cdot \sum_{i=1}^n (z_i(A))^2 \geq 1 + 2 \cdot \sqrt{2/n}$.

In particular, for $n = 7$, we mark a point A as a fault if $(1/n) \cdot \sum_{i=1}^n (z_i(A))^2 \geq 2.07$. The result is slightly better than union, average, or DS, but edges are still marked as fault, and also, the resulting picture contains a lot of separate points erroneously marked as faults.

7. OUR SECOND IDEA: DISMISSING NON-CONFIRMED MEASUREMENT RESULTS

The reason why separate points are marked as outliers is that one of the fused measurement results may be erroneous (caused by sensor malfunctions). An erroneous measurement result $x_i(A)$ will, most probably, be drastically different from the mean a_i . As a result, the corresponding value $|z_i(A)|$ will be large and the value $(z_i(A))^2$ will also be large. So, we will erroneously mark this point as an fault.

To avoid this situation, we should not mark the point A as a fault if only one measurement results in an abnormal value $x_i(A)$. We would rather only consider a point as a fault, if at least two different measurements detect an outlier there.

Thus, we arrived at the following modified method for combining the values of $z_i(A)$ for each point A :

- If at this point,
 - only one of the values $z_i(A)$, $1 \leq i \leq n$, indicates a fault (i.e., $|z_i(A)| > 2$), and
 - in a neighborhood of a certain radius d , no other measurement detects any fault (i.e., $|z_j(B)| \leq 2$ for all $j \neq i$ and for all points B for which $d(A, B) \leq d$),

then we dismiss this large value of $z_i(A)$, and only combine the remaining values $z_j(A)$, $j \neq i$.

- For all other points A , we combine all the values $z_i(A)$, $1 \leq i \leq n$.

This idea drastically decreases the number of separate points marked as faults.

8. OUR THIRD IDEA: DETECTING FAULTS ON EDGES

How can we now decrease the number of false positives (FP) on edges? The reason for these FP is that for fault detection, a typical fault may be a hole (or a thinning) in the plate, i.e., in effect, an extra edge.

To avoid showing edges as faults, we modified the above algorithm as follows: for points near the edges, we compared the measured characteristic $x_i(A)$ not with the average value a_i and standard deviation σ_i over the entire plate, but only with the average $a_i^{(e)}$ and standard deviation $\sigma_i^{(e)}$ taken over the edge points. Then, for edge points A , we compute the normalized values $z_i(A)$ as $z_i(A) = (x_i(A) - a_i^{(e)})/\sigma_i^{(e)}$, and apply the above data fusion techniques.

9. RESULTING METHOD

In summary, our proposed fault detection method consists of the following steps:

- first, we apply the iterative method to the entire plate and to the edge zone only, and get values a_i , σ_i , $a_i^{(e)}$, and $\sigma_i^{(e)}$; based on these values, we compute the normalized values $z_i(A)$;
- we dismiss large values $z_i(A)$ that are not confirmed by other measurements;
- then, for each point A , we combine the values $z_i(A)$ and declare A a fault if $\sum(z_i(A))^2$ exceeds the corresponding threshold.



Figure 1. New Statistical Method

10. COMPARISON WITH THE PREVIOUS METHODS

In our sample plate, we considered all the points within $3/4''$ from the edge as edge points. The new method did indeed drastically decrease the number of FN and FP, to the extent that we could see that there was a minor error in the originally referencing of the images in comparison with the original plate: the reconstructed fault locations are slightly shifted with respect to the actual ones. So, to make a fair comparison of different methods, we first shift the resulting pictures to eliminate the reconstruction shifts.

We select the shift for which the total number of FN and FP is the smallest. For all the methods, the best shift is $s = 3$. For this shift, with our method, we have 11 false negatives and 5986 false positives. To compare with our DS method, we choose a threshold for which $FN=11$; for this threshold, the new method has 10281 false positives.

For comparison with union and average, the results are even better. In other words:

*For the same amount of false negatives,
our new method leads to an almost 50% decrease
in the number of false positives.*

One might argue that a possible reason for the success of the new method is that perhaps we tuned this method to perform on the test plate only. To check whether the new method indeed better in general than the OH method, we tested it on other test plates.

These two $250 \text{ mm} \times 250 \text{ mm}$ test plates are manufactured by placing 24 thin layers of metal on top of each other and subjecting the plates to high pressure. We had two plates with 12 faults on each plate, 4 faults arranged systematically on each of the 3 rows: three circles of diameters $1''$, three circles of diameters $1/2''$, three circles of diameters $1/4''$, and three amorphous faults.

Three circular faults from the first row are created by placing Teflon circles on the 6th layer, the second row on the 12th layer and the third row on the 18th layer. The amorphous faults are created by placing balloons of diameter 1". The upper fault was placed at the 6th layer, the second at 12th layer, the third on 18th layer. Under pressure, balloons change shape and circular faults shift.

An ultrasonic signal was sent to each plate, and the reflected pulse echo signals were detected at several different moments of time. Different times of arrival of pulse echo correspond to 8 different depths.

To test our algorithm, we applied this algorithm and our version of DS method to both new plates. Since we do not know the exact locations of the faults, we cannot directly compare the quality of fault detection of these two methods. So, we used the following comparison technique. As already mentioned, because the faults are slightly shifted, we do not know where exactly they are. For the amorphous faults, we do not know the actual shape either, so, we only analyze circular faults. We know the 9 zones, each of which contains one circular fault of given radius; within each of these zones, we found a circle of given radius which best fits the set of points marked as faults in this zone – i.e., the circle for which the number of un-marked points within this circle is the smallest.

We consider this circle as an (approximate) location of the actual fault, and we count false positives and false negatives with respect to this location.

Similarly to the original test plate, to compare the new methods with other methods, we changed the parameters of the other methods so that they lead to the same number of false negatives as our new method, and compared the number of false positives.

For the first new plate, our method led to 59 false negatives and 1792 false positives; for the same number of FN, the modified DS method led to 16272 FP – almost 10 times more (and even more for average and union).

For the second new plate, our method led to 32 false negatives and 1315 false positives; for the same number of FN, the modified DS method led to 24193 FP – almost 20 times more (and even more for average and union). For both plates, the new method works much better.

11. TOWARDS FURTHER IMPROVEMENT: USING CORRELATION

To get a better fault detection, we can look into the assumptions that we made. One such assumption is the statistical independence of different measurements.

Our statistical analysis shows that although some pairs of measurements are indeed almost independent, there is some correlation between some other measurement results. For example, there is a sizable correlation between two measurements RES 2 and RES 3 of resonance corresponding to different frequencies: the corresponding correlation coefficient is $\approx 54\%$.

For correlated normally distributed random variables, the probability is a function not of $\sum_{i=1}^n (z_i(A))^2$, but of a more general quadratic form $\sum_{i=1}^n \sum_{j=1}^n b_{ij} \cdot z_i(A) \cdot z_j(A)$, where the matrix b_{ij} is the inverse to the covariance matrix C_{ij} that describes the correlation between the measurements. Each element C_{ij} of this matrix is the average of the product $z_i(A) \cdot z_j(A)$ over different points A . It is therefore reasonable to do the following:

- compute the elements of the covariance matrix C_{ij} ;
- invert this matrix getting b_{ij} ; and then
- use the expression $\sum_{i=1}^n \sum_{j=1}^n b_{ij} \cdot z_i(A) \cdot z_j(A)$ instead of the expression $\sum_{i=1}^n (z_i(A))^2$ to detect the faults.

For the Boeing benchmark plate, in the edge zone, the distribution was not very homogeneous in the sense that when we applied our algorithm to different parts of the edge zone, we got drastically different covariance values C_{ij} . Due to this non-homogeneity, we only applied the covariance method to the interior zone.

In the interior zone, we have an improvement, albeit a small one:

- instead of 1 false negative, we have 0 false negatives, and
- instead of 1,892 false positives, we have 1,862 false positives.

12. WE ARE ALSO INTERESTED IN THE SHAPE OF A FAULT

After the faults are detected and located, we need to make a decision on whether these faults are dangerous or not. Whether a fault is dangerous depends on its shape: From mechanics, we know that the angular shapes are subjected to more stress than the smooth ones; therefore, if the shape of the fault is angular, then this fault is more dangerous than a smooth one. Thus, we need to know not only the locations of the faults, but also their shapes.

The above methods reproduce the fault locations well, but the shapes of the reconstructed faults do not always match the shapes of the actual faults: some of the square (dangerous) faults are reproduced as circular (less dangerous) ones. How can we modify the above method so that it will better reproduce the shapes of the faults?

The main reason why the above method is not very good in detecting shape is that in the above method, our main objective was not to miss any faults – because faults are dangerous. Therefore, when there was good evidence to support both hypotheses: that the pixel A is a fault and that the pixel A is not a fault – we tended to declare it a fault. As a result, we “padded” the set of fault points with extra points – thus distorting the shape of the set of all the fault points.

To get the shape better, it is therefore reasonable to treat the two hypotheses equally. Specifically, we consider two hypotheses: H_0 that a point is not a fault and H_1 that the point is a fault, and we use the standard techniques of hypothesis testing (see, e.g., Wadsworth¹⁵) to decide which of these hypotheses is more probable: we choose H_1 if the ratio p_1/p_0 of the probability of H_1 to the probability of H_0 exceeds a certain threshold t .

We already know how to describe the probability of H_0 . It turns out that the distribution of $z_i(A)$ for fault points is also approximately Gaussian – of course, with different values a_i^f and σ_i^f . (We can estimate the values a_i^f and σ_i^f by processing the points that are marked as outliers after the iterative algorithm for computing a_i and σ_i .) As a result, after taking logarithms of both sides, we can transform the criterion $p_1/p_0 \geq t$ to an equivalent form

$$\sum_{i=1}^n \left((z_i(A))^2 - \left(\frac{z_i(A) - a_i^f}{\sigma_i^f} \right)^2 \right) \geq t_0.$$

Same arguments as before lead us to choose the same value for t_0 .

Important point: to apply this algorithm, we need to have enough fault point to be able to conclusively estimate a_i^f and σ_i^f ; there were enough such point in the interior of the plate, but not enough on the edge; so, we only applied this method in the interior.

As a result, we got a better shape reconstruction. Specifically, to estimate the quality of the shape reconstruction of a square fault, we took the set F of all the points marked as faults in the neighborhood of an actual fault, found the square S that is the closest to the resulting set (in the sense that the total number of pixels in the symmetric set difference is the smallest), and then estimates the number of false positives and false negatives by comparing F and S . Similarly, for circular faults, we compare the set F with the closest circle C . Then, we take the total number of false positives over all 8 inside faults, and the total number of false negatives over these faults. Here is the result of our comparison:

- When we apply this procedure to the original method, we got 2,443 false positives and 19 false negatives inside the plate.
- For the new method, we got 1,895 false positives and 11 false negatives inside the plate.

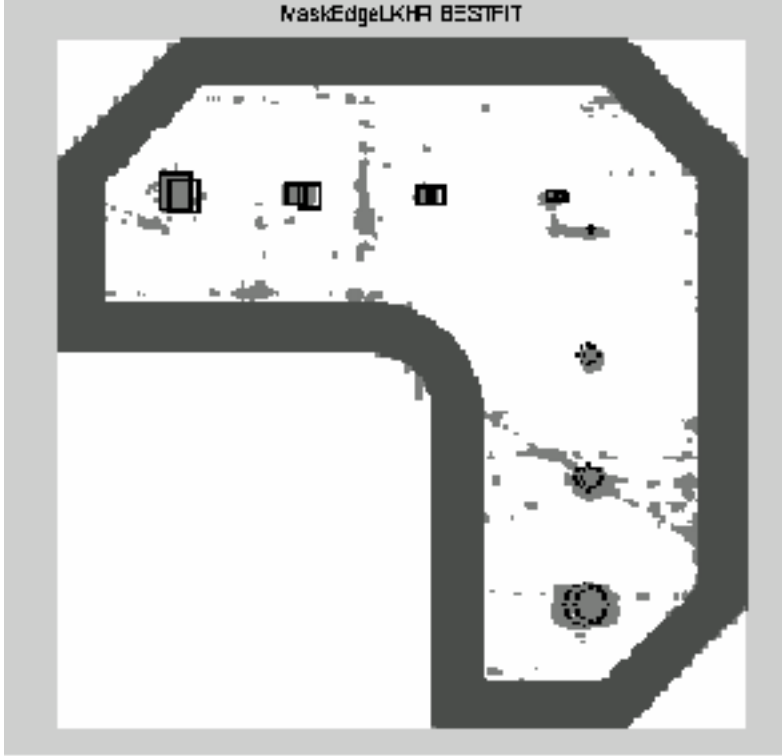


Figure 2. A Method for Reconstructing Shapes

Conclusion: the new method is indeed much better in reconstructing shapes.

13. CONCLUSIONS

Several techniques are known for non-destructive testing of aerospace structures, such as pulse echo, Eddy current, magnetic resonance, etc. Each of these techniques detects some faults but misses others, so it is desirable to combine (fuse) the results of these techniques. Several methods of data fusion are known; we tested them on a benchmark sample: a plate from B-52 with artificially induced faults of different type and location. The best fault detection came from the simplest possible method - union, where a point is declared a fault if the result of one of its measurements differs from the norm. Alas, union leaves several fault points undetected.

To improve the quality of fault detection, we modified the straightforward statistical method as follows:

- we computed mean and variance iteratively: detected outliers are excluded from the computation on the next iteration;
- we treated the plate's edge and the inside separately;
- we dismissed measurements in which only one technique detects an outlier as possibly erroneous.

The resulting method indeed leads to a much better detection. We also successfully tested our method on other examples of composites with artificial fault at different depths. Taking correlation into consideration leads to a slight further improvement.

Once a fault is located, it is important to decide how dangerous is this fault; to make this decision, we must know the fault's size and the fault's shape. In this paper, we propose a modification of the above method that is specifically tailored towards reconstructing these shapes.

ACKNOWLEDGMENT

This work was supported in part by the Future Aerospace Science and Technology Program (FAST) Center for Structural Integrity of Aerospace Systems, effort sponsored by the Air Force Office of Scientific Research, Air Force Materiel Command, USAF, under grants numbers F49620-95-1-0518 and F49620-00-1-0365. It was also supported by the Air Force Research Laboratory, Air Force Material Command, USAF, under grant number F33615-99-C-5211. The US Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright notation thereon.

The work was also partly sponsored by NASA under cooperative agreement NCC5-209 and grant NCC2-1232, and by NSF grants CDA-9522207, EAR-0112968, EAR-0225670, and 9710940 Mexico/Conacyt.

REFERENCES

1. C. Ferregut, R. A. Osegueda, Y. Mendoza, V. Kreinovich, and T. J. Ross, "Aircraft Integrity and Reliability", In: *Combined Fuzzy Logic and Probability Applications*, J. Booker, J. Parkinson, and T. J. Ross (eds.), Combined Fuzzy Logic and Probability Applications, pp. 219–242, SIAM, Philadelphia, 2002.
2. C. Ferregut, R. A. Osegueda, and A. Nuñez (eds.), *Proc. International Workshop on Intelligent NDE Sciences for Aging and Futuristic Aircraft*, El Paso, TX, September 30–October 2, 1997.
3. X. E. Gros, *NDT Data Fusion*, J. Wiley, London, 1997.
4. A. C. Holguin, "Fusion of NDE Data for Inspecting Composite Patches on Aging Aircraft", *Abstracts of The Model Institutions of Excellence 6th Annual Conference*, El Paso, TX, April 4–8, 2001, p. 37.
5. G. Klir and B. Yuan, *Fuzzy Sets and Fuzzy Logic: Theory and Applications*, Prentice Hall, Upper Saddle River, NJ, 1995.
6. M. Krishna, V. Kreinovich, and R. A. Osegueda, "Fuzzy Logic in Non-Destructive Testing of Aerospace Structures", *Proc. 1999 IEEE Midwest Symposium on Circuits and Systems*, Las Cruces, New Mexico, August 8–11, 1999, Vol. 1, pp. 431–434.
7. B. C. Mulupuru, *Differentiating between angular and smooth shapes in noisy computer images, on the example of non-destructive testing of aerospace structures*, Master's Thesis, University of Texas at El Paso, 2002.
8. H. T. Nguyen and E. A. Walker, *First Course in Fuzzy Logic*, CRC Press, Boca Raton, FL, 1999.
9. R. A. Osegueda, C. Ferregut, V. Kreinovich, S. Seelam, and H. Schulte, "Fuzzy (Granular) Levels of Quality, With Applications to Data Mining and to Structural Integrity of Aerospace Structures", *Proc. 19th International Conference of the North American Fuzzy Information Society NAFIPS'2000*, Atlanta, Georgia, July 13–15, pp. 348–352, 2000.
10. R. Osegueda, V. Kreinovich, L. Potluri, and R. Aló, "Non-Destructive Testing of Aerospace Structures: Granularity and Data Mining Approach", *Proc. FUZZ-IEEE'2002*, Honolulu, Hawaii, May 12–17, Vol. 1, pp. 685–689, 2002.
11. R. A. Osegueda, A. Revilla, L. Pereyra, and O. Moguel, "Fusion of modal strain energy differences for localization of damage", In: A. K. Mal (ed.), *Nondestructive Evaluation of Aging Aircraft, Airports, and Aerospace Hardware III*, *Proc. SPIE*, 1999, Vol. 3586, Paper 3586-28.
12. R. A. Osegueda, S. R. Seelam, A. C. Holguin, V. Kreinovich, and C.-W. Tao, "Statistical and Dempster-Shafer Techniques in Testing Structural Integrity of Aerospace Structures", *International Journal of Uncertainty, Fuzziness, Knowledge-Based Systems (IJUFKS)*, **9**(6), pp. 749–758, 2001.
13. L. Potluri, *Correlation techniques for information fusion*, Master's Thesis, Univ. Texas at El Paso, 2002.

14. S. R. Seelam, *Intelligent outlier detection and information fusion methods for noisy images, with applications to structural integrity of aerospace structures*, Master's Thesis, Univ. Texas at El Paso, 2002.
15. H. M. Wadsworth, Jr. (editor), *Handbook of statistical methods for engineers and scientists*, McGraw-Hill Publishing Co., N.Y., 1990.