Statistical and Dempster-Shafer Techniques in Testing Structural Integrity of Aerospace Structures

Roberto A. Osegueda
Seetharami R. Seelam
Ana C. Holguin
Vladik Kreinovich

FAST Center
University of Texas
El Paso, TX 79968, USA
vladik@cs.utep.edu

Chin-Wang Tao

Department of
Electrical Engineering
National I-Lan
Institute of Technology
260 I-Lan, Taiwan
cwtao@mail.ilantech.edu.tw

Abstract

We describe the existing statistics-related methods of testing structural integrity for aerospace structures, describe their drawbacks, how they can be overcome, and compare the resulting techniques.

Keywords: structural integrity, aerospace structures, statistical methods, Dempster-Shafer method

1 Introduction

Testing of structural integrity of aerospace structures is an important practical problem (see, e.g., [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. Each of these techniques detects some faults but misses others, so it is desirable to combine (fuse) the results of these techniques.

In this paper, we describe different statistical and Dempster-Shafer methods for producing such a combination, describe the optimal variant of these techniques, and test all these methods on a real-life example.

2 Test Study

To compare different fault detection and fault localization techniques, we have taken a plate from an actual airplane (B-52) and artificially added several faults to this plate. In total, we added 16 faults: 8 squares and 8 circles. Of these 16 faults, 8 are inside the plate:

- four squares of sides $1/2''$, $1/3''$, $1/4''$, and $1/8''$; and
- four circles of diameters $1/2''$, $1/3''$, $1/4''$, and $1/8''$;

and 8 are of the exact same size but on the edge of the plate:

- four squares of sides $1/2''$, $1/3''$, $1/4''$, and $1/8''$; and
- four circles of diameters $1/2''$, $1/3''$, $1/4''$, and $1/8''$.

For this plate, we have 7 different measurements. None of these methods could detect all the faults by itself without "marking" non-fault points as faults. To be specific, we had the following results:

- Pulse Echo 0 (PE0) well detected inside square faults, and somewhat fuzzily all the others.
- Pulse Echo 1 (PE1) detected the boundaries of all the faults, but not what is inside these boundaries.
- The Magnitude of Eddy Current (EC) measurements detect only the boundary of the largest inside circular fault. Since the EC magnitude does not seem to contribute to fault detection, we decided not to combine it with other measurements.
- The Eddy Current phase (EC0) detected all inside circular faults and nothing else.
- The magnitude of Magnetic Resonance measurements 0, 1, 2, and 3 did not detect any faults at all

(RES2 fuzzily detected the largest circular fault inside), so we decided not to combine them with other measurements.

- The phase of Magnetic Resonance 0 (RES0) detected all circular faults near the plate's edge.
- The phases of Magnetic Resonance 1 and 2 (RES1, RES2) also detected (but more fuzzily) all circular faults near the plate's edge.
- The phase of Magnetic Resonance 3 (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 particularly, practically all of these methods marked several points on the edge of the plate as faults.

3 Traditional Statistical Approach

The *general idea* of non-destructive fault detection is as follows:

- If one of these techniques leads to the unusual signal value (very different from the normal values), then a fault is probably located at this point.
- The larger the difference, the larger is the fault.

To implement this idea, for each measurement, we compute:

- the (arithmetic) *average* a of all the values, and
- the mean square (standard) deviation σ of the signal x from its average a.

In accordance with the standard statistical techniques (see, e.g., [9]), we can then consider values x which are outside of the "two sigma" interval $[a - 2\sigma, a + 2\sigma]$ as possible faults. In other words, we compute the "normalized" values $z = (x - a)/\sigma$, and consider points for which $|z| > 2$ to be possible fault locations.

4 The Main Drawback of the Traditional Statistical Approach

The problem with this approach is that the used estimates for average and standard deviation are *biased* because they use not only regular values, but fault values as well.

The computation of the the average a and of the standard deviation σ involve also the fault points, in which the value of the signal is very different from the expected normal value. Hence, the computed σ is much larger than for the normal points. As a result, some true faults – which would have been detected by a correct (smaller) σ – go undetected when we use this larger value.

In our example, for some measurements, the values of x corresponding to most actual fault locations are within a σ or 2σ from the average. Thus, the resulting points are not therefore recognized as fault locations.

5 Iterative Approach: Our Idea

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

- first, we use the traditional procedure and find some faults;
- then, we re-calculate a and σ by omitting known faults;
- we then check for new faults by using these better estimates for a and σ,

and we continue these iterations until no new fault locations are uncovered (it usually takes 3-4 iterations).

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

- new approximations $a^{(s)}$ and $\sigma^{(s)}$ are computed for the desired values of a and σ; 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 $a^{(s)}$ and the standard deviation $\sigma^{(s)}$ based on all un-marked points of the plate;
- then, we check each un-marked point, and mark all the points for which $|x - a^{(s)}| > 2\sigma^{(s)}$.

The process stops when the subsequent values of $a^{(s)}$ and $\sigma^{(s)}$ become close to each other (this usually happens on the 3rd or 4th steps). The corresponding final values of $a^{(s)}$ and $\sigma^{(s)}$ are then taken as the desired values of a and σ.

As a result, for each measurement and for each point, we get a normalized value $z = (x - a)/\sigma$.

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.

6 How To Combine The Normalized Measurement Values

Since for each measurement i, the distribution for non-fault points is Gaussian with 0 mean and unit standard deviation, the probability that a point for which the normalized measurement result is z_i does not have a fault is proportional to $\exp\left(-z_i^2/2\right)$.

Since different measurement detect different types of faults, it is natural to assume that different measurements are statistically independent. Thus, the probability (density) that a non-fault point can have the measured values $z_1, \ldots, z_n$ is equal to the product of the probabilities corresponding to different measurements, and is, therefore, proportional to

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

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 $\sum z_i^2 \geq \text{const}_2$, or, equivalently,

$$\frac{\sum_{i=1}^{n} z_i^2}{n} \geq \text{const}_3,$$

where n is the total number of measurements. 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 $\bar{z}$ of the values z_i^2 is distributed approximately according to the normal distribution. The mean $E[\bar{z}]$ of this distribution is equal to the mathematical expectation $E[z_i^2]$ of z_i^2, i.e., to $E[z^2] = \sigma[z] = 1$. The mean square $E[(\bar{z} - E(\bar{z}))^2] = \sigma^2[\bar{z}]$ of the difference between the average $\bar{z}$ 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., [9]). Thus, the average $\bar{z}$ has the mean 1 and the standard deviation $\sqrt{2/n}$. Thus, if $\bar{z} \geq 1 + 2 \cdot \sqrt{2/n}$, then we can be sure that there is fault at this point.

So, the criterion for declaring a point a fault location is

$$\frac{\sum_{i=1}^{n} z_i^2}{n} \geq 1 + 2 \cdot \sqrt{\frac{2}{n}}.$$

For $n = 7$, we thus get $\text{const}_3 = 2.07$; thus, the criterion is equivalent to

$$\sqrt{\frac{\sum_{i=1}^{n} z_i^2}{n}} \geq \sqrt{2} \approx 1.4.$$

If we want only to consider certain faults, then it may be better to use the "three sigma" rule instead of the "two sigma" rule. In this case, we get $\text{const}_3 = 1 + 3 \cdot \sqrt{2/n} = 2.6$, so the criterion is equivalent to

$$\sqrt{\frac{\sum_{i=1}^{n} z_i^2}{n}} \geq \sqrt{2.6} \approx 1.6.$$

7 How To Improve the Combination Rule

As we have mentioned, one of the measurements may be erroneous, so we would not rather consider a point a fault location if at least two different measurements detect a fault there. On the other hand, we do not want to just dismiss one of the measurements, because if two different measurements point to the same place as a fault location, we thus have more confidence that this is indeed a fault location. Thus, we used the following idea of combining the values of $p_i(A)$ 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., is $z_i(A) > 2$), and
  - in a neighborhood of a certian 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 $z_i(A)$, $1 \leq i \leq n$.

8 Detecting Faults on Edges

The resulting method detects faults inside the plate, but is still has trouble distinguishing between faults near the plate edge and the regular edge points.

The reason for this is that for all the above testing methods, unusual values occur not only at the fault locations, but also at the edges of the plate. This relation between faults and edges is very natural, since, e.g., a typical fault is a hole (or a thinning) in the plate, i.e., 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 not with the average value a and standard deviation σ over the entire sample, but only with the average a_e and standard deviation σ_e taken over the edge points. Then, we compute the normalized value as $z = (x - a_e)/\sigma_e$. As a result, we got a much better description of the faults near the edges.

Instead of dividing the plate into two zones: points which are close to the edge, and points which are inside the plate, we used an even finer subdivision into three zones:

- points which are very close to the edge (distance from the edge between 0 and $1/4''$);
- points which are somewhat close to the edge (distance from the edge between $1/4''$ and $1/2''$);
- points which are inside the plate (distance from the edge larger than $1/2''$).

For each of these three zones, we computed separate values of a and σ and used these values to normalized the measurement results in the corresponding zone. In this manner, we got an even better description of the faults near the edges.

9 Statistical Approach: Results

The combination of the above new ideas indeed leads to a successful detection of the faults on our testbed
B-52 plate.

10 Future Work: How To Further Improve the Combination Rule

The fault detection is good but not perfect. To get a better fault detection, we can look into the assumption that we made. One such assumption is the statistical independence of different measurements. Our statistical analysis shows that although the measurements are indeed almost independent, there is is some correlation between them. For correlated normally distributed random variables, the probability is a function not of $\sum\limits_{i=1}^{n} z_i^2$, but of a more general quadratic form $\sum\limits_{i=1}^{n}\sum\limits_{j=1}^{n} b_{ij}\cdot z_i\cdot z_j$.

We therefore plan to do the following:

- compute the correlation matrix C_{ij} as the average of the product $z_i(A)\cdot z_j(A)$ for different points A;
- invert this matrix getting b_{ij}; and then
- use the expression $\sum\limits_{i=1}^{n}\sum\limits_{j=1}^{n} b_{ij}\cdot z_i\cdot z_j$ instead of the expression $\sum\limits_{i=1}^{n} z_i^2$ to detect the faults.

11 Dempster-Shafer Approach to Non-Destructive Testing

In this paper, we have described a new idea on how to improve the traditional statistical approach to non-destructive testing. Researchers have also used another idea: using Dempster-Shafer formalism instead of the traditional statistical techniques.

In this approach, for each i-th measurement, we first transform the normalized measurement results $z_i(A)$ in different points A into probabilities $p_i(A)$ by using the follow-up "normalization" into the interval $[0,1]$ of possible values of probability:

- $p_i(A)=0$ if $z_i(A)<0$, and
- $p_i(A):=z_i(A)/Z_i$ if $z_i(A)\ge 0$, where
$$Z_i=\max_A z_i(A).$$

This value $p_i(A)$ is interpreted as the probability that, according to i-th measurement, there is a fault at the point A. We can then, assuming that these measurements are independent, combine these probabilities in a different way:

- We can compute the *plausibility* $Pl(A)$ that there is a fault at the point A as the probability that at least one of the measurements detected a fault. This probability can be computed as follows:
  - the probability that, according to i-th measurement, there is no fault at the point A, is equal to $1-p_i(A)$;
  - since the measurements are assumed to be statistically independent, the probability that there is no fault according to all the measurements is equal to the product $\prod_i(1-p_i(A))$;
  - finally, the probability that there is a fault 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 $Bel(A)=\prod_i p_i(A)$.

We can then outline the fault location as points in which, e.g., one of the these probabilities exceeds a certain threshold.

In spite of the seemingly *ad hoc* nature of this method, it leads to a pretty reasonable fault localization.

12 Why Do Dempster-Shafer Methods Work: A Problem

The "probabilities" used in this approach differ from the actual frequencies; thus, it is not clear why this method, which relies on these crude approximations for probabilities, leads to such good fault localization results.

13 Why Do Dempster-Shafer Methods Work: An Explanation

We have an explanation for this success, and this explanation is as follows. Let us first present this explanation on the example of the belief function $Bel = \prod p_i$. If we use belief, then the criterion for a point to be a fault is to have $\prod_i p_i \geq \text{const}_1$ for some threshold constant const_1. Substituting the expression for probabilities in terms of measurement values ($p_i = z_i/Z_i$) into this formula, we get an equivalent criterion $\prod_i z_i \geq \text{const}_2$ for some new constant const_2. By turning to logarithms, we conclude that this criterion is equivalent to $\sum_i \ln(z_i) \geq \text{const}_3$.

We know that in non-fault points, the values z_i are distributed according to normal distribution with 0 average and unit standard deviation, and we know that these measurements are independent. Thus, for a large number of measurements, we can say that the sum $\sum_i \ln(z_i)$ is normally distributed, with a certain average a_s and standard deviation σ_s. Hence, we can conclude that if the actual value of the sum is larger than $a_s + 2 \cdot \sigma_s$, then the corresponding point is most probably a fault. This explains why the criterion $\sum_i \ln(z_i) \geq \text{const}_3$ makes sense.

Similarly, for plausibility, the criterion

$$1 - \prod_i (1 - p_i) \geq \text{const}_1$$

is equivalent to

$$\prod_i (1 - p_i) \leq \text{const}_2$$

hence to

$$\sum_i \left(-\log \left(1 - \frac{z_i}{Z_i} \right) \right) \geq \text{const}_3.$$

This sum-using criterion can be statistically justified in the same way as we have justified the similar sum-criterion for the belief function.

14 The Existing Dempster-Shafer Methods Behave Slightly Worse Than The New Statistical Method

We have shown that for serious faults, both belief- and plausibility-based criteria detect the faults.

For smaller faults, the detection may not be that sure, because the probabilities that we use in this method are different from the actual ones. Not surprisingly, our experimental results showed that the existing Dempster-Shafer methods behave slightly worse than the new statistical method: Namely, even for the best of these methods (based on plausibility), both the numbers of false positives (points erroneously classified as faults) and false negatives (un-detected fault points) are smaller for the new statistical method.

15 Future Work: Improving the Existing Dempster-Shafer Techniques

One possible reason why the existing Dempster-Shafer techniques do not work so well is that they rely on the estimates of when either at least one measurement detects a fault or all measurement detect the fault. We have already mentioned that:

- one measurement may be too little, because in some locations, one measurement erroneously locates a fault;
- on the other hand, waiting for all the measurements to detect a fault is too much, because, as we have mentioned, some faults are overlooked by several measurements (this is the main reason why we need to combine several measurements in the first place).

So, we plan to improve the existing Dempster-Shafer method by considering not just when at least one or when all measurements detect the fault, but also when, e.g., at least two measurements detect the fault.

We can reformulate this idea in interval terms.

- In the Dempster-Shafer approach, we use an *interval* to describe the uncertainty instead of an exact value as in probabilistic approach.
- In this approach, traditionally, the interval between belief and plausibility is used, where, crudely speaking, belief describes to what extent all measurements predict a fault, while plausibility describes to what extent at least one of the measurements predicts a fault.
- For our purposes, this interval is too wide.
- To narrow this interval, we propose to use modified measures which describe, e.g., to what extent at least two measurements predict a fault.

How can we compute these new measures? The probability that none of the measurements detected the fault is $P_0 = \prod(1 - p_i)$. The probability that exactly one measurement has detected the fault is equal to

$$P_1 = \sum_i p_i \cdot \prod_{j \neq i} (1 - p_i) = P_0 \cdot \sum_i o_i,$$

where

$$o_i \stackrel{\text{def}}{=} \frac{p_i}{1 - p_i}.$$

Similarly, the probability that exactly 2 measurements detected the fault is equal to

$$P_2 = P_0 \cdot \sum_{i<j} o_i \cdot o_j,$$

the probability that exactly k measurements detected the fault is equal to

$$P_k = P_0 \cdot \sum_{i_1 < \ldots < i_k} o_{i_1} \cdot \ldots \cdot o_{i_k},$$

and the probability that at least k measurements detected the fault is equal to $1 - P_0 - P_1 - \ldots - P_{k-1}$.

Computing P_0 and P_1 requires $O(n)$ computational steps, where n is the total number of measurements. If we try to compute P_2 by directly using the above formula, we will need $O(n^2)$ computational steps. A faster way to compute this expression is to use the fact that

$$(\sum_i o_i)^2 = \sum_i o_i^2 + 2 \sum_{i<j} o_i \cdot o_j,$$

and therefore,

$$\sum_{i<j} o_i \cdot o_j = \frac{1}{2} \cdot \left((\sum_i o_i)^2 - \sum_i o_i^2 \right).$$

This formula requires $O(n)$ steps.

Similarly, we can get $O(n)$ formula for computing P_3 as an appropriate linear combination of the expressions $\left(\sum_i o_i\right)^3$, $\left(\sum o_i\right)^2 \cdot (\sum_i o_i)$, and $\sum_i o_i^3$. Other combinations lead to P_4, P_5, etc.

Our preliminary experiments show that this modification indeed leads to a better fault detection.

16 Conclusions

In this paper, we:

- Developed a new method which improves the current statistics-based outlier-detecting methodology of locating potential faults by using iterative outlier detection.
- Developed a new method for combining fault probabilities coming from different measurements, method which better takes into consideration that some of the measurements may be not very reliable.
- Developed a new method for separating measurement abnormalities caused by actual faults from abnormalities caused by the closeness to the edge, thus improving the quality of fault detection.
- Combined the above new methods into a new fault-detection methodology and successfully tested it on the example of simulated B-52 faults.

For Dempster-Shafer techniques, we:

- Explained, from the statistical viewpoint, why alternative Dempster-Shafer techniques work so successfully in fault detection.
- Developed an improved method of computing belief and plausibility in Dempster-Shafer approach which leads to a better fault detection.

Acknowledgment

This work was supported in part by NASA under cooperative agreement NCC5-209, by NSF grants No. DUE-9750858 and CDA-9522207, by the United Space Alliance, grant No. NAS 9-20000 (PWO C0C67713A6), 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 grant number F49620-95-1-0518, and by the National Security Agency under Grant No. MDA904-98-1-0561.

The authors are thankful to Hung T. Nguyen for valuable discussions.

References

[1] Ferregut, C., Osegueda, R. A., Mendoza, Y., Kreinovich, V., and Ross, T. J., "Aircraft Integrity and Reliability", In: Booker, J., Parkinson, J., and Ross, T. J. (eds.), *Combined Fuzzy Logic and Probability Applications*, SIAM, 2000 (to appear).

[2] Ferregut, C., Osegueda, R. A., and Nuñez, A. (eds.), *Proceedings of the International Workshop on Intelligent NDE Sciences for Aging and Futuristic Aircraft*, El Paso, TX, September 30–October 2, 1997.

[3] Gros, X. E. *NDT Data Fusion*, J. Wiley, London, 1997.

[4] Klir, G., and Yuan, B., *Fuzzy Sets and Fuzzy Logic: Theory and Applications*, Prentice Hall, Upper Saddle River, NJ, 1995.

[5] Krishna, M., Kreinovich, V., and Osegueda, R. A., "Fuzzy Logic in Non-Destructive Testing of Aerospace Structures", *Proceedings of the 1999 IEEE Midwest Symposium on Circuits and Systems*, Las Cruces, New Mexico, August 8–11, 1999, Vol. 1, pp. 431–434.

[6] Nguyen, H. T., and Walker, E. A., *First Course in Fuzzy Logic*, CRC Press, Boca Raton, FL, 1999.

[7] Osegueda, R. A., Ferregut, C., Kreinovich, V., Seelam, S., and Schulte, H., "Fuzzy (Granular) Levels of Quality, With Applications to Data Mining and to Structural Integrity of Aerospace Structures", *Proceedings of the 19th International Conference of the North American Fuzzy Information Society NAFIPS'2000*, Atlanta, Georgia, July 13–15, 2000, pp. 348–352.

[8] Osegueda, R. A., Revilla, A., Pereyra, L., and Moguel, O., "Fusion of modal strain energy differences for localization of damage", In: Mal, A. K. (ed.), Nondestructive Evaluation of Aging Aircraft, Airports, and Aerospace Hardware III, Proceedings of SPIE, 1999, Vol. 3586, Paper 3586-28.

[9] Wadsworth, H. M., Jr. (editor), *Handbook of statistical methods for engineers and scientists*, McGraw-Hill Publishing Co., N.Y., 1990.