

University of Wales Trinity Saint David
School of Engineering

Structural Integrity and Material Evaluation
M7X01384

Zachary Church
1700054

Contents

Figures.....	iii
1 Introduction	1
2 Literature Review.....	2
2.1 Introduction	2
2.2 POD	3
2.3 Dimensioning	5
3 Statistical Analysis.....	7
3.1 Statistical Metrics	7
3.2 Distributions.....	8
3.3 Sizing Plots	10
3.4 Statistical Conclusion	14
4 Conclusion.....	15
5 References	16

Figures

Figure 1. Typical POD Curve.....	4
Figure 2. Ideal POD Curve.	4
Figure 3. Operator 9 Sizing Chart.....	6
Figure 4. Provided Sizing Data.	7
Figure 5. Mean, Standard Deviation, 95% Probability.....	7
Figure 6. All Normal Distributions.	8
Figure 7. Operator 8 Distribution.	9
Figure 8. Operator 7 Distribution.	10
Figure 9. Operator 4 Sizing.	11
Figure 10. Operator 1 Sizing.	11
Figure 11. operator 9 Sizing.....	12
Figure 12. Operator 2 Sizing	13
Figure 13. All Operator Sizing.	13

1 Introduction

Nondestructive testing (NDT) methods are not totally accurate, which means they carry a probability of detection (POD) and are also not 100% accurate when sizing flaws, cracks, and wall thicknesses. The development of assessing the probability of detecting flaws started with NASA and their development of a program for assessing NDT methods for the development of various space programs. Sizing accuracy is a concern when assessing NDT methods, statistical methods can be used to analyze the performance of a selected NDT method and the operators. The operators of nondestructive testing procedure can have a huge effect of the accuracy of testing. For this paper a set of data was provide for a statistical analysis and assessment of a through wall thickness sizing procedure and an assessment on the 10 operators defined in the data.

2 Literature Review

2.1 Introduction

A litany of non-destructive testing methods exists and have been developed from the 20th century and continue to be developed. NDT testing methods range from x-ray methods, magnetic, visual, and Ultrasonics. All the different types of ways to detect flaws in materials are nothing without answering the question of how small can be and still be detected. More realistically the question should be what is the largest flaw that will not be picked up by that particular testing method. (Rummel. & Matzkanin, 1997) All NDT methods have different POD curves because of many factors these include operator skill, surface preparation, material, and difficulty of administering the tests. Certain methods of nondestructive testing methods are better suited to different applications and would have a higher chance of detection a certain flaw sizes Ultrasonic testing methods are very versatile and can be used across many different applications and many different size applications.

Published in 1997 The “Nondestructive Evaluation (NDE) Capabilities Data Book (Third Edition)” summarizes the standards and sets forth guidelines from engineers to the technicians that carry out the testing methods. (Rummel. & Matzkanin, 1997) The document was funded by the United States department of defense, so standards and data published in this document at least partially stemmed from NASA. In 1969, NASA saw the need for nondestructive testing on space craft, because of the harsh environmental factors that space craft go through NDT methods are important for spacecraft and aeronautical industries. This was done for the development of the space shuttle program. (National Aeronautics and Space Council , 1969) NASA developed a program to develop the POD detection methods that are used today. This was later adopted by the United States air force and the aero space industry as a whole. NASA used a program to derive the Probability of detection curves for that method of testing. To do this they created a controlled experiment by manufacturing cracks and prepping the material for a number of operators to run a test and detect the manufactured crack. (RUMMEL, 2010)

2.2 POD

The Probability of detection is not a reflection of the reliability of the nondestructive testing method being used. The calculated POD is merely a prediction of the flaws that will be picked up with a certain testing method. To predict the flaws that might be picked up a controlled experiment like the one NASA used. (RUMMEL, 2010) There are four possibilities that can happen when operators run tests during controlled experiments. These are shown in the table below. (Matzkanin & Yolken, 2001)

True Positive	There is a flaw, and it is detected.
False Positive	A flaw is detected, and a flaw not Present.
False Negative	A Flaw is present, but it is not detected.
True Negative	A flaw is not detected, and a flaw is not present.

Table 1. Detection Possibilities.

The operator of the specific testing method records the flaws that they might be finding, that is then compared to a master list of known flaws in a material. probability of detection can then be calculated by Equation 1.

Equation 1. Probability of Detection.

$$PoD = \frac{True\ Positives}{True\ Positives + False\ Negatives}$$

The probability of false alarm can also be calculated by Equation 2. This would be used to determine the odds of the chosen NDT method has to register a flaw or crack that might not be there. This could be as troublesome as not finding a crack at all.

Equation 2. Probability of False Alarm.

$$PoFA = \frac{False\ Positives}{True\ negatives + False\ Positives}$$

Probability of detection can best be conveyed as a confidence percentage. For example above a certain flaw size you can be 80% or 95% confident that a flaw will be detected if it is there. For most applications it would be useful to be 95% confident to detect a flaw that would be large enough to cause the part to be deemed unusable and need replacement. (Matzkanin & Yolken, 2001)

Error! Reference source not found. shows a typical POD curve, this conveys the idea that as the flaw size increases the probability of detection increases and starts moving to 100%. POD Curves are

created from Hit/Miss data where there is some cross over where some operators would find the flaw and some operators would miss it giving the curve the shape it has. The Ideal POD curve would look like the green line in **Error! Reference source not found.** (George A Georgiou, Jacobi Consulting Limited, 2006)

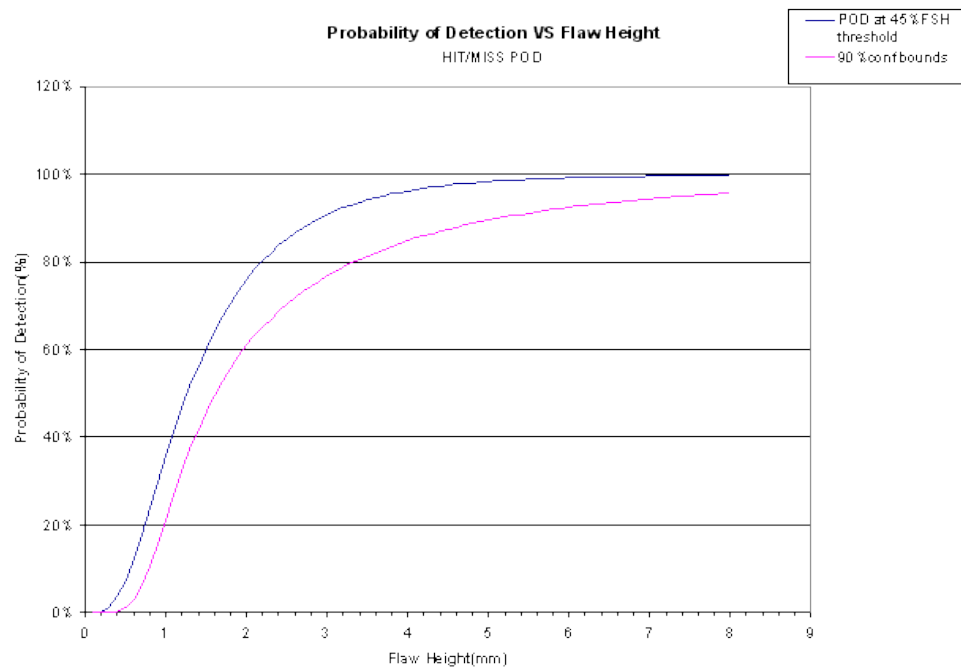


Figure 1. Typical POD Curve. (Ginzel, 2006)

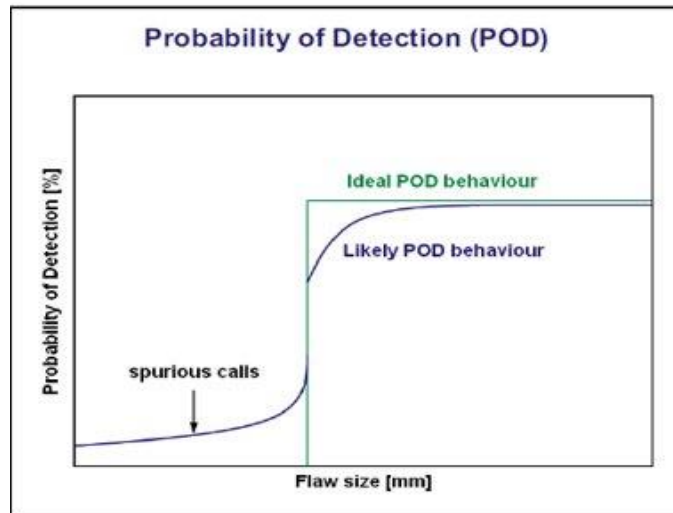


Figure 2. Ideal POD Curve. (GRANVILLE & CHARLTON, n.d.)

2.3 Dimensioning

Dimensioning flaws is also important to the field on nondestructive testing. To know how large a deformity in a part is, allows the manufacturer to accurately life cycle the parts they have put into production. In the oil and gas sector, storage containers need to be monitored because of a corrosion risk to said storage contains, failure of the storage containers pose a great environmental and public safety threat. The aeronautical industry needs to keep strict records and closely monitor airplanes because airplanes suffer from fatigue cracks forming, NASA was a big proponent of NDT testing in the aeronautical industry because of the aforementioned space shuttle program and developing a program to derive the probability of detection. Under strict FAA rules airlines need to constantly monitor their aircraft for cracks stemming from riveted joints in the structure of the aircraft. Accidents like the Aloha Airlines flight 243 in 1988 highlighted the need for monitoring fatigue cracks in the fuselage of airplanes. (Airways, 2020) Once the cracks reach a critical point that component needs to be replaced or the airplane retired. The best way to monitor Fatigue cracks in airplanes is with modern NDT methods allowing any detected cracks to be dimensioned. Dimensioning cracks too big would cost airlines billions in unneeded costs as retire airplanes early would be very expensive. At the same time dimensioning a crack far too small could potentially cause a material failure creating the potential for an explosive decompression event like the Aloha tragedy. Different methods of NDT have different ways of dimensioning detected flaws. Along with the fact that each NDT method has a different way of dimensioning flaws operators may make mistakes in dimensioning or detection flaws. NDT operators

need to go through extensive training and be certified by a governing body to ensure that they are using the correct processes while operating NDT equipment. Some NDT Operators might be fundamentally better than others as it is with every job, Training methods can allow for some confidence that NDT operators are doing the right thing and skilled enough.

Sizing accuracy can be analyzed statistically to examine the accuracy of an NDT method over a range of Operators this would allow the organization running the tests to examine operator skill and the overall effectiveness of the specified NDT method. Tests much like the aforementioned NASA tests can be run to produce data on how accurately flaws can be located and sized. With known flaws in a material and a known size for each Operators can size the flaws themselves. The recorded data can then be organized into a graph like the one below.



Figure 3. Operator 9 Sizing Chart.

A line can be drawn marking the true size of the flaws then plotted against that a trend line can be plotted using the recorded flaw sizes. This allows for the user to plot trendlines to visualize the spread of data from the true sizes. Lines at ± 0.5 mm, ± 1 mm and so on can be plotted to visualize where most of the data falls in. In the example above the operator tends to be sizing systematically 0.5mm larger than the true size. A less scattered distribution of data across multiple operators would point to an NDT method that could potentially reliably size flaws that are detected in the field. What constitutes a flaw needs to be decided on before sizing can begin. In reality every little imperfection could be potentially considered a flaw. Flaw sizing comes with errors because the sizing relies heavily on

the operator. The Gaussian or Normal distributions can be used to estimate how an operator could sized certain flaws. (HSE, 2008)

3 Statistical Analysis

3.1 Introduction

For this paper a statistical analysis a set of data was provided to assess new ultrasonic thickness gauging procedure, see Figure 4. The data provided included 21 thickness measurements by ten different operators.

Defect	true size	Op1	Op2	Op3	Op4	Op5	Op6	Op7	Op8	Op9	Op10
A	0.25	0.3	0.5	0.0	0.5	0.0	1.0	0.5	1.0	0.8	0.5
B	0.27	1.2	1.2	0.5	1.0	0.0	1.0	0.5	1.0	0.8	0.7
C	0.28	0.7	1.0	1.2	0.5	0.0	1.0	0.5	1.0	1.3	0.8
D	0.4	0.8	0.5	1.0	1.0	1.0	1.2	0.6	1.0	1.5	0.9
E	0.4	0.9	1.2	1.0	1.1	1.0	1.4	0.2	1.0	1.6	1.0
F	0.5	0.5	0.9	1.0	0.8	1.0	1.1	0.0	1.2	1.0	0.8
G	0.52	0.7	1.5	1.1	1.0	1.5	0.9	1.0	1.0	0.9	1.0
H	0.55	0.4	1.2	0.8	1.2	0.9	0.5	1.3	1.0	1.6	1.0
I	0.7	1.0	1.5	0.9	1.2	1.6	2.0	1.6	1.0	2.2	1.4
J	0.98	1.4	1.0	0.8	1.6	2.0	1.9	0.5	1.0	1.8	1.3
K	1.12	1.3	0.5	1.0	1.8	1.8	1.0	1.8	1.5	2.0	1.4
L	1.15	1.2	1.8	1.3	1.5	1.0	2.2	1.8	1.5	2.0	1.6
M	1.1	1.6	2.0	1.4	1.5	1.6	2.0	0.5	2.0	2.2	1.6
N	1.3	1.7	2.1	0.9	2.0	2.0	2.2	1.7	1.5	2.0	1.7
O	1.6	1.5	0.8	0.9	1.5	2.0	2.0	1.2	2.0	2.0	1.6
P	1.8	1.1	2.0	1.4	3.2	2.0	2.4	2.2	2.5	2.5	2.1
Q	2.1	2.6	1.8	2.3	3.0	2.5	3.5	2.0	2.0	3.2	2.5
R	2.7	2.2	3.0	3.2	3.5	2.5	2.8	2.5	3.5	3.5	3.0
S	4.9	5.5	4.5	4.5	4.4	4.5	5.7	4.2	5.5	5.4	4.9
T	6	6.0	5.5	6.2	5.5	6.5	7.0	5.3	6.5	6.7	6.2
U	7.5	6.5	6.0	6.3	6.5	7.2	8.0	6.5	8.0	6.9	7.0

Figure 4. Provided Sizing Data.

	Op1	Op2	Op3	Op4	Op5	Op6	Op7	Op8	Op9	Op10
Mean	0.1	0.2	0.1	0.4	0.3	0.7	0.0	0.5	0.8	0.3
Standard Dev	0.45591248	0.676109	0.513859	0.540356	0.460568	0.396987	0.551283	0.254332	0.426172	0.25837
95% Probability	0.88902934	1.318412	1.002026	1.053694	0.898108	0.774126	1.075003	0.495947	0.831036	0.503821

Figure 5. Mean, Standard Deviation, 95% Probability.

The mean was calculated by taking difference of each measurement taken by each operator to the true size given with the data, then taking the average of the differences, this essentially gives the mean of the error for each of the operators. The standard deviation was then calculated using the standard deviation function in Microsoft Excel. The range of data used to calculate the standard deviation was again the difference in each measurement from the true size. Neither of these values give a good picture of how an operator or a process is performing. Operators 1 and 3 received the same mean scores but sized flaws differently and received different standard deviation scores and different 95% probability scores.

The 95% probability of the size of the flaw can communicate how accurate a particular operator has been. The number calculated is the amount of error in millimeters on either side of the true size that the operator could be being 95% confident. For example Operator 7 had a 95% probability of 1.07mm either side of the true size of any given measurement. Compared to operator 7, Operator 8 had a 95% probability of 0.495, this points to operator 8 being more accurate than operator 7. The 95% probability was calculated using the following equation

Equation 3. 95% Probability.

$$95\% \text{ Probability} = \text{Standard Deviation} \times 1.95$$

3.2 Distributions

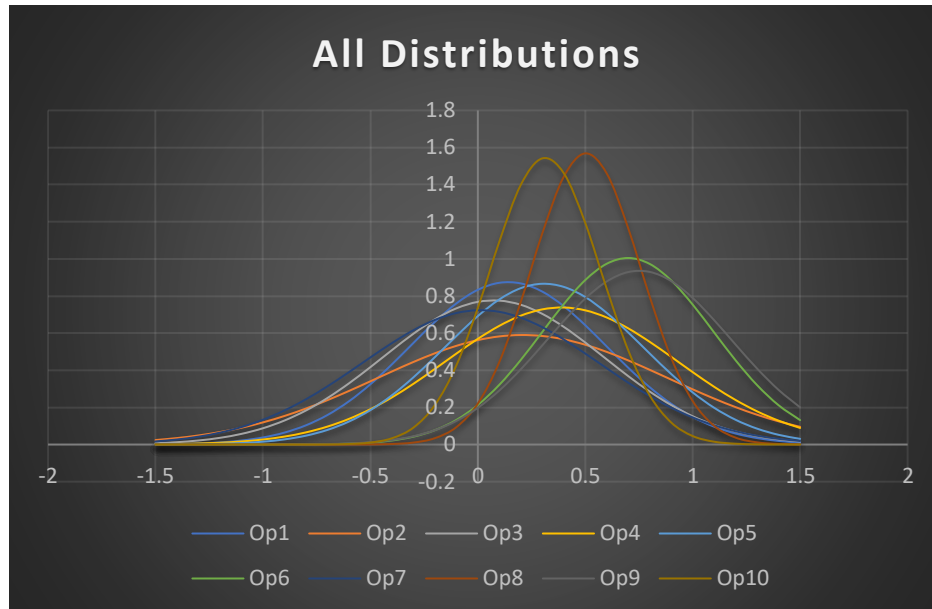


Figure 6. All Normal Distributions.

The normal distribution was calculated over a range of values from -1.5 to 1.5, this was done by using the normal distribution function in Microsoft Excel. The normal distribution graphs relate to the mean and standard deviation. The peak of the curve is where the mean of the data lies, 68.2% of the data lies within 1 standard deviation of the mean. Each normal distribution graph visually looks skewed, but the means of the data sets just lie above zero moving the curve to the right of zero. Shown below are two different normal distribution graphs from operators 7 and 8. Operator 8 had a higher standard deviation while operator 7 had a lower standard deviation. Having a smaller standard deviation forces the data set to a small range meaning that the operator might be a little more accurate. Operator 7 has a wider range of values creating a wide distribution of data.

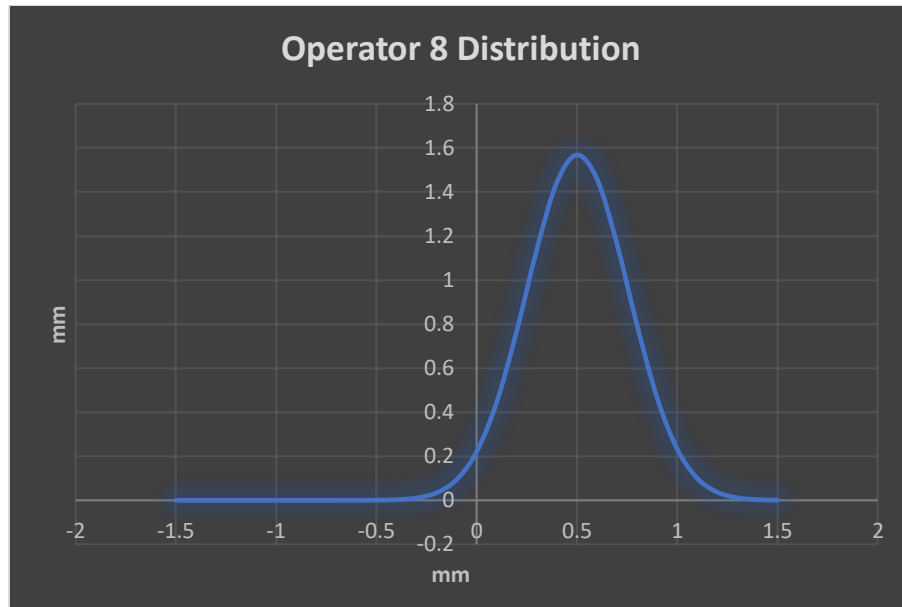


Figure 7. Operator 8 Distribution.

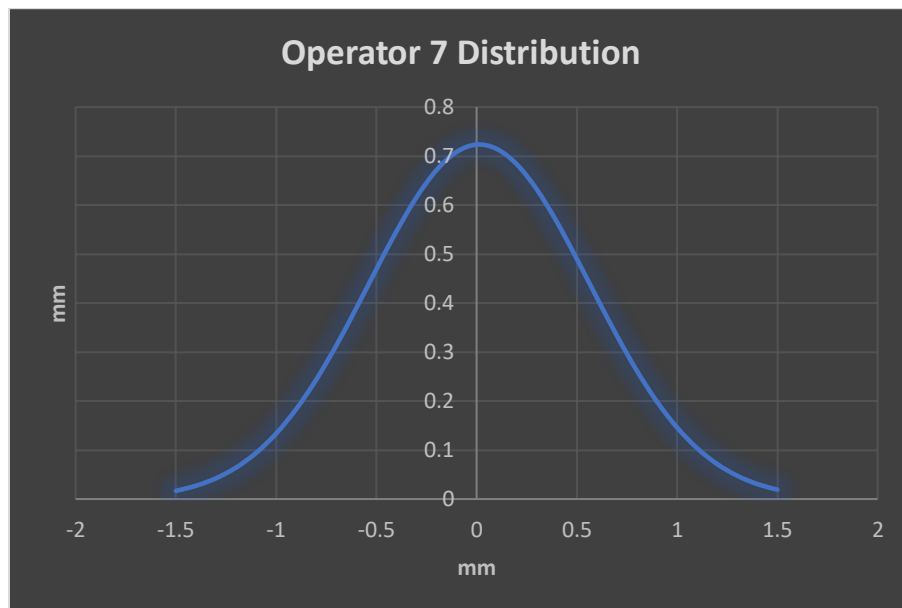


Figure 8. Operator 7 Distribution.

3.3 Sizing Plots

Dimensioning through wall thickness will have some variation no matter what, with each of the ten operators there are variation in measurements between each of them. The though wall thickness measurements can be organized into the graphs shown below. Trendlines can be plotted over the data

to show how the operator might size through wall thicknesses relative to the true size. Operator 1 was shown to be sizing through wall thicknesses large when the true size was small but sizing the through wall thickness small when the true size is larger. The trendline would indicate that the larger the through wall thickness the more likely they are to measure it too small. This is also shown in operator 4's graph, operator 4 was shown to do the same as operator 1 but more exaggerated. Operator four would size smaller flaws much bigger than that of operator 1 and much smaller than operator 1 at the higher through wall thicknesses.



Figure 9. Operator 4 Sizing.

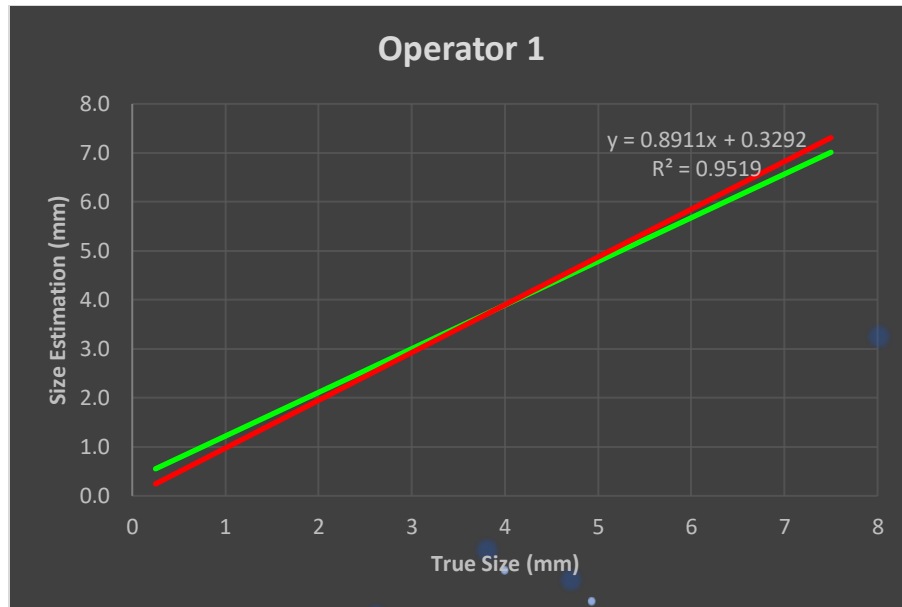


Figure 10. Operator 1 Sizing.

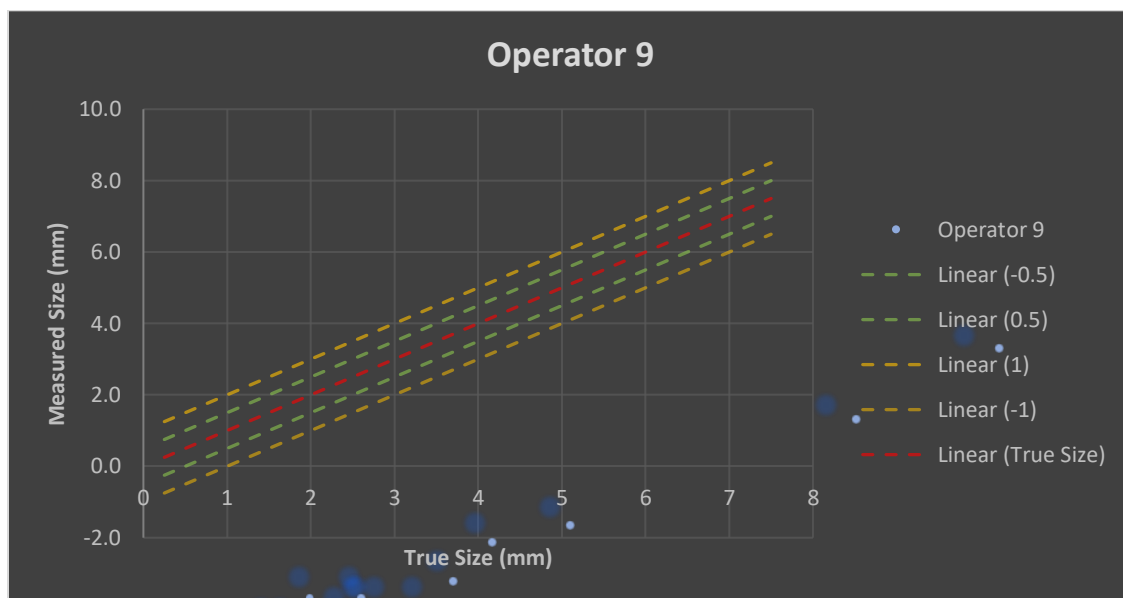


Figure 11. operator 9 Sizing.

Data can also be displayed in the format above to get an idea of how a specific operator is sizing the through wall thickness. The lines on the graph show where the measurements would be if they we

0.5mm larger and smaller and 1mm larger and smaller. Figure 11 shows that operator 9 consistently at least 0.5mm larger than the true size especially with the smaller thicknesses. Operator 9 has one data point at the largest through wall thickness that is lower than the true size, this falls in line with the fact that the smaller through wall thicknesses will get sized larger than their true size while the larger through wall thicknesses would get sized smaller than their true size.

Operator 2, Figure 12, was shown to have a larger distribution of measurements but still had the same trend of overestimating through wall thickness size but still might underestimate the sizes at the larger thicknesses.

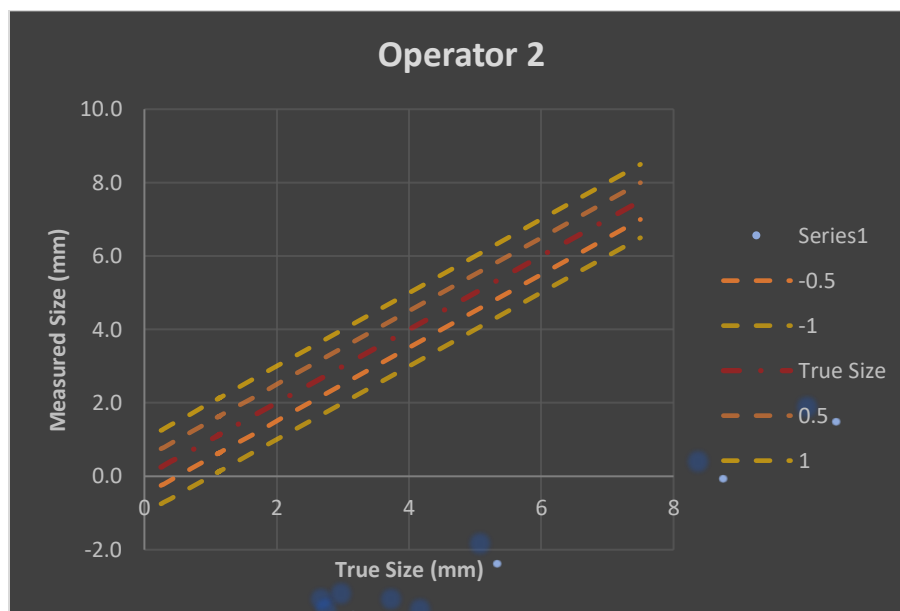


Figure 12. Operator 2 Sizing

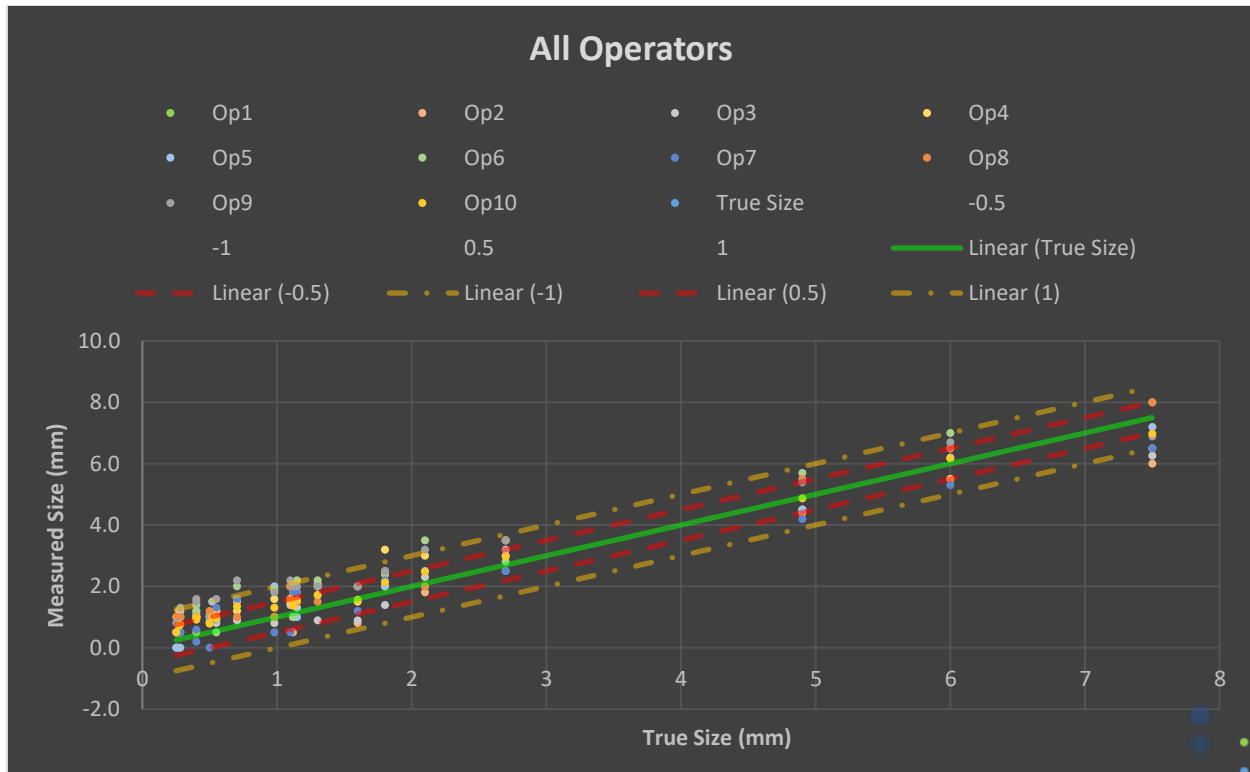


Figure 13. All Operator Sizing.

Figure 13 shows the same data that is shown in Figure 11 and Figure 12 but compiled from every one of the operators. Visually this shows that most of the data at the smaller true sizes is overestimated by all of the operators. At the larger thicknesses the data is more split on either side of the true size, but for the largest wall thickness the operators tended to underestimate the thickness.

3.4 Statistical Conclusion

Examining the data given for statistical analysis conclusions about the ultrasonic through wall testing methods can be drawn. Examining normal distributions, sizing charts and the standard statistical metrics like mean and standard deviation we can see that this particular method of through wall sizing tends to overestimate the size of the thickness no matter what operator was at the controls until the wall thickness gets above somewhere around 4.5mm. If this error is found to be consistent a correction factor could be applied. The through wall thicknesses of 2.7mm to 6mm had the best distribution of data either side of the true size.

4 Conclusion

The development of NDT methods leads to the necessitation of developing ways to assess how different forms of NDT perform and how operators might affect their performance. In the 1970's NASA developed a program to create a method to assess the reliability of NDT methods. Each method has its own POD and can be affected by the operator, material, and flaw type. Sizing accuracy is another way to assess the reliability of NDT methods. Sizing accuracy is important because the size of the flaw, crack, or wall thickness dictates when parts, materials, or machines, meet the end of their lifecycle. Ending a part or machines life cycle early could cost a company a huge sum of money so accurately dimensioning a crack or wall thickness is important. Data was provided for this paper for a statistical analysis on the performance of the NDT method for through wall sizing. It was found that this particular sizing method was found to size smaller wall thickness too large and larger wall thickness too small, this was the case for all 10 of the operators that completed the test. The optimal range for this particular testing method might be in the range of 2.5mm to 6mm.

5 References

George A Georgiou, Jacobi Consulting Limited, 2006. *Probability of Detection (PoD) curves, Derivation, applications and limitations*, London: Health and Safety Executive.

Airways, 2020. *Airways Magazine*. [Online]

Available at: <https://airwaysmag.com/airlines/32-years-alo-ha-flight-243-accident/>

[Accessed 24 12 2021].

Annis, C., L. G. & Martin, O., n.d. *Optimal Sample Size for Probability of Detection Curves*, Palm Beach Gardens, Florida; Petten, The Netherlands: Statistical Engineering; European Commission, Joint Research Centre, Institute for Energy and Transport.

Ginzel, E., 2006. *ndt.net*. [Online]

Available at: <https://www.ndt.net/article/v11n05/ginzel2/ginzel2.htm>

[Accessed 27 12 2021].

GRANVILLE, R. K. & CHARLTON, P. C., n.d. *Detection and Sizing of Baffle Plate Erosion and Fretting using Eddy Current Array Technology*, Swansea: UWTSD.

HSE, 2008. *INFORMATION FOR THE PROCUREMENT AND CONDUCT OF NDT, Part 4: Ultrasonic Sizing Errors and Their Implication for Defect Assessment*, s.l.: s.n.

Matzkanin, G. A. & Yolken, H. T., 2001. *A Technology Assessment of Probability of Detection (POD) for Nondestructive Evaluation (NDE)*, Austin, Texas: NTIAC.

National Aeronautics and Space Council , 1969. *Aeronautics and Space Report of the President*, Washington D.C.: Executive Office of the President .

Rummel, W. D. & Matzkanin, G. A., 1997. *Nondestructive Evaluation (NDE) Capabilities Data Book*. Third ed. Austin, Texas: NTIAC.

RUMMEL, W. D., 2010. *NONDESTRUCTIVE INSPECTION RELIABILITY - HISTORY, STATUS AND FUTURE PATH*, Littleton, Colorado: D&W Enterprises, LTD..