

Analysis of Operators Performing a New Ultrasonic Detection Method

Kyle Turner

January 2022

1 Introduction

Non-Destructive Testing (NDT) or Non-Destructive Evaluation methods have been used throughout human history, far before the term Non-Destructive Testing was used. Humans have used visual inspection to search for surface imperfections and cracks, listened to metal as it was worked and tested the sound after it is cast. Modern Non-Destructive Testing methods have evolved to use more sophisticated technology such as radiography and acoustic emission but old methods have still remained in use due to their efficacy. An Non-Destructive Testing method is assessed by its capability, reproducibility and repeatability [1], with the Probability of Detection (PoD) being the current standard to measure of its capability. This report shall discuss the history and development non-destructive testing of probability of detection and sizing methods and will perform a statistical analysis of a new ultrasonic thickness gauging procedure in order to assess the sizing accuracy of the individual technicians, as well as the equipment and procedure as a whole.

2 Probability of Detection

The equation for probability of detection is defined by NASA as the statistical estimate of the proportion of all flaws of a given size that will be detected in a particular NDE inspection [2]. This can be calculated using $PoD = \frac{\text{total number of positive calls}}{\text{total opportunities for rejection}}$, where the total positive calls is the number of true positives detected and the total opportunities for rejection is the sum of the true positives and false negatives. Generally this value will be calculated for each size of defect and plotted on a probability of detection curve such as that shown in Figure 1 to create a continuous distribution of results allowing you to estimate how likely an Non-Destructive Testing method is to give you an accurate answer. Furthermore using this information you can generate an additional fit to display the 95% confidence band, providing a visual representation of the defect size at which you have 95% confidence that a defect will be detected.

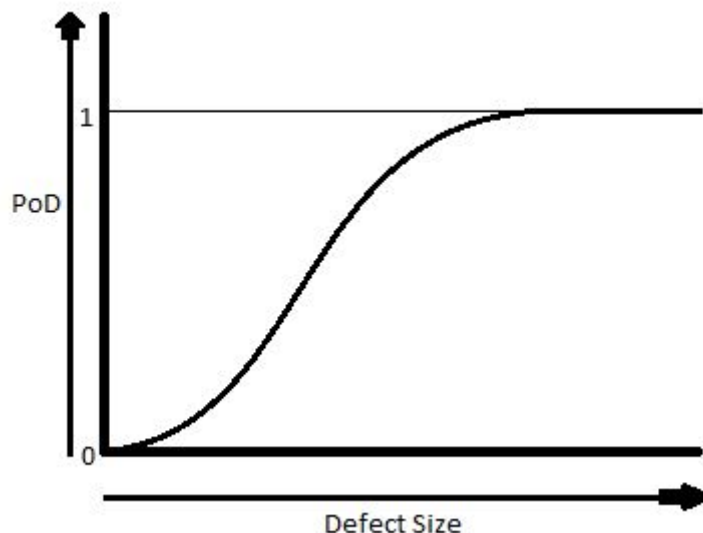


Figure 1: An example PoD curve.

Probability of Detection as a method was developed in in 1969 National Aeronautics and Space Administration (NASA) started a program to to determine the largest possible flaw size which could be missed by an NDT method, very soon the United States Air Force US commercial aircraft industries adopted these standards. The method to procure a probability of distribution curve required the procurement of a number of known flaw samples of different sizes and types, use an NDT method to detect the flaws and record them as a function of their flaw size. This can then be plotted as a PoD curve. The available methods for generating a PoD curve uses either Hit/Miss data or signal

response, a lot more information is transmitted back from the signal response data [3]. The PoD curve for hit/miss data is written as:

$$PoD(a) = \frac{e^{\frac{\pi}{\sqrt{3}} \frac{\ln a - m}{\delta}}}{1 - e^{\frac{\pi}{\sqrt{3}} \frac{\ln a - m}{\delta}}} \quad (1)$$

Where m is the median and δ is the standard deviation. A minimum of 60 flaws, evenly distributed between the maximum and minimum flaws of interest is recommended for hit/miss data.

For signal/response data, due to the increased data recieved, less flaws are required to achieve the same accuracy and so only 30 samples are recommended, however having more samples will provide a more accurate PoD curve. The PoD curve for signal/reponse data is derived from the relationship between a and $\hat{a}$. It has been observed in a number of studies that a roughly linear relationship exists between $\ln a$ and $\ln \hat{a}$ and the relationship is usually expressed as:

$$\ln \hat{a} = \alpha_1 + \beta_1 \ln(a) + \gamma \quad (2)$$

Where γ is an error term and $\alpha_1 + \beta_1$ is the mean of the probability density function of $\ln \hat{a}$. In this data a flaw is considered detected if it exceeds some threshold $\hat{a}_{thr}$. The PoD function can thus be expressed as:

$$PoD(a) = Probability(\ln \hat{a} > \ln \hat{a}_{thr}) \quad (3)$$

3 Non-Destructive Testing Sizing Methods

Of the sizing techniques used in NDT as ultrasonic techniques are the most common and of most interest to the latter half of this paper they shall be concentrated on in this section. There are four main methods of ultrasonic sizing; dB drop technique, phased array and time of flight diffraction (TOFD). The use of ultrasonics in NDT started with a proposal by Soviet physicist S.Y Sokolov who proposed the through-transmission technique for detecting flaws in metals [4]. During World War 2 many nations independently developed their own pulse-echo methods and come 1945 the first commercially available pulse-echo instruments became available.

Amplitude based techniques were the first method used [5], wherein the amplitude of the returning signal correlates to the flaw size, a calibration curve using the area of known flat bottom holes and signal amplitude is created and used to estimate flaw size. The distance to the flaws is easily calculated using the expression:

$$d = \frac{vt}{2} \quad (4)$$

Where d is the distance from the surface to the defect, v is the velocity of the sound wave in the material and t is the time between emission and receiving the signal. This technique only works for flaws with a diameter less than the beam size.

The time of flight diffraction method is a method developed to inspect welds as it allowed the inspection of the weld surfaces. TOFD uses two transducers which, assuming the diffracting tip is centered allows the depth of the tip to be calculated using the following equation:

$$Depth = \sqrt{\frac{BPL^2}{2} - \frac{HD^2}{2}} \quad (5)$$

Where BPL is the beam path length for the signal and HD is the beam path length for the lateral wave. This is limited in its capability to detect near surface flaws when the arrival time difference between lateral and longitudinal signal is similar to the pulse length of the signal.

The 6dB and 20dB drop method uses a single probe being swept across a defect, as the probe is swept across the defect the received by the probe is reduced. 6dB corresponds to the amplitude dropping by 50 percent and so the distance where the amplitude drops to 50 percent at each edge of the flaw is considered the length of the flaw. As the 6dB drop method can undersize the flaw you can instead use the 20dB drop method which corresponds to 10 percent of the signal strength, however this suffers issues with signal to noise due to how small the signal is.

The phased array setup consists of a series of probes which, by timing when each probe fires, can be used to steer the wave or to concentrate the wave on a location. This allows the technician to optimise the beam angle to get the best signal as well as allowing a much larger volume of material to be scanned from a single position, however due to the larger size of the array it is more difficult to guarantee good coupling with the inspected material and equipment is typically more expensive than a traditional setup [6]. This is further limited by the necessity that the flaw be within the near-field of the probe as it will oversize flaws beyond it.

4 Statistical Analysis

In order to assess the sizing accuracy of the individual technicians and the new ultrasonic thickness gauging method and equipment as a whole, ten technicians have used the equipment and method in order to size 21 known defects. This information has been collected and presented in Figure 2. Using Microsoft Excel's Analysis Toolpak to process this information I have started by obtaining the mean, standard deviation and standard error of each technician and each defect as this will allow me to judge the accuracy of the method as a whole and for each individual technician.

Defect	true size	Op1	Op2	Op3	Op4	Op5	Op6	Op7	Op8	Op9	Op10
A	0.25	0.3	0.5	0	0.5	0	1	0.5	1	0.8	0.5
B	0.27	1.2	1.2	0.5	1	0	1	0.5	1	0.8	0.7
C	0.28	0.7	1	1.2	0.5	0	1	0.5	1	1.3	0.8
D	0.4	0.8	0.5	1	1	1	1.2	0.6	1	1.5	0.9
E	0.4	0.9	1.2	1	1.1	1	1.4	0.2	1	1.6	1
F	0.5	0.5	0.9	1	0.8	1	1.1	0	1.2	1	0.8
G	0.52	0.7	1.5	1.1	1	1.5	0.9	1	1	0.9	1
H	0.55	0.4	1.2	0.8	1.2	0.9	0.5	1.3	1	1.6	1
I	0.7	1	1.5	0.9	1.2	1.6	2	1.6	1	2.2	1.4
J	0.98	1.4	1	0.8	1.6	2	1.9	0.5	1	1.8	1.3
K	1.12	1.3	0.5	1	1.8	1.8	1	1.8	1.5	2	1.4
L	1.15	1.2	1.8	1.3	1.5	1	2.2	1.8	1.5	2	1.6
M	1.1	1.6	2	1.4	1.5	1.6	2	0.5	2	2.2	1.6
N	1.3	1.7	2.1	0.9	2	2	2.2	1.7	1.5	2	1.7
O	1.6	1.5	0.8	0.9	1.5	2	2	1.2	2	2	1.6
P	1.8	1.1	2	1.4	3.2	2	2.4	2.2	2.5	2.5	2.1
Q	2.1	2.6	1.8	2.3	3	2.5	3.5	2	2	3.2	2.5
R	2.7	2.2	3	3.2	3.5	2.5	2.8	2.5	3.5	3.5	3
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	5.5	6.2	5.5	6.5	7	5.3	6.5	6.7	6.2
U	7.5	6.5	6	6.3	6.5	7.2	8	6.5	8	6.9	7

Figure 2: Data collected from ten technicians using the new ultrasonic thickness gauging method.

Using the descriptive statistics function of Excel I have collected the mean size for each defect and plotted it with the true size and the technicians sizes in Figure 4

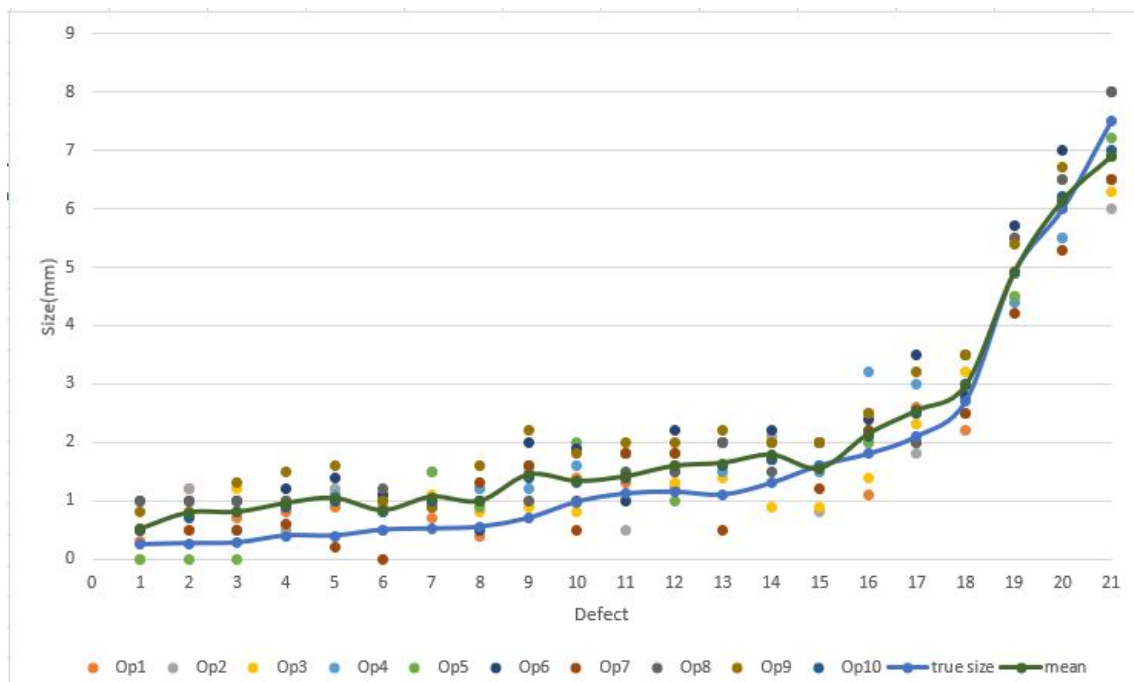


Figure 3:

Looking at this table we can see for the first 14 defects are on average being oversized by the technicians, this however changes from the 18th defect onward where they are quite similar. In order to better observe the difference between the mean size collected by the technicians and the true size I have calculated the standard error and plotted the mean size with error bars in Figure ?? with the true size.

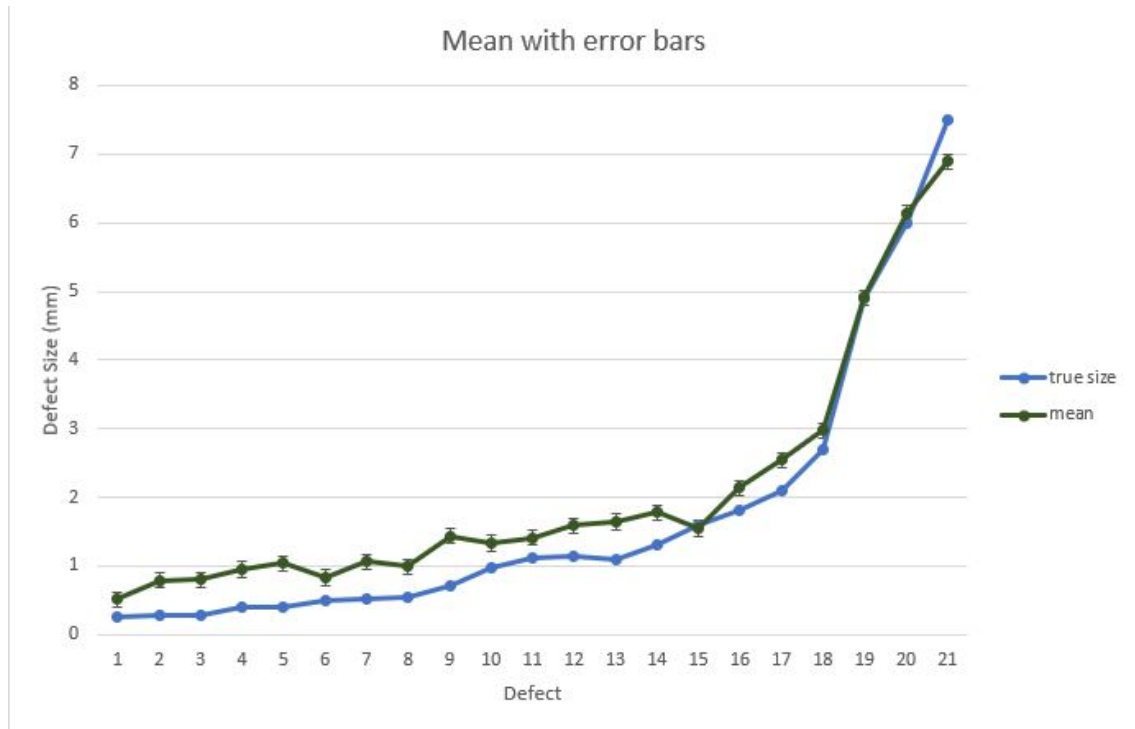


Figure 4: A plot showing the mean size plotted with the true size

From this we can clearly see that for most results the mean size detected by the technicians lies outside of the error bars with defect A, M, R, S and T being very close. It is noteworthy that R, S and T are all larger size defects which may imply that the method does not work well for smaller defect sizes and tends to oversize them. It is also possible that the equipment has a zero error and tends to undersize larger defects however this is difficult to test as we have significantly less large defects to test this theory on. The final solution I will propose is that human error, the technicians inexperience with the method and equipment, has resulted in the mistakes during the smaller results as they would be more difficult to locate and size accurately, however this would be unlikely to occur to all technicians.

In order to test the theory that there is perhaps a zero error in the equipment I have calculated the difference between the mean size detected by the technicians and true size, I have then found the mean of these values. This has been calculated to be 0.3419. I will take this away from the mean sizes calculated by the technicians and plot it on a graph against the true size to see how well it fits visually in Figure 5

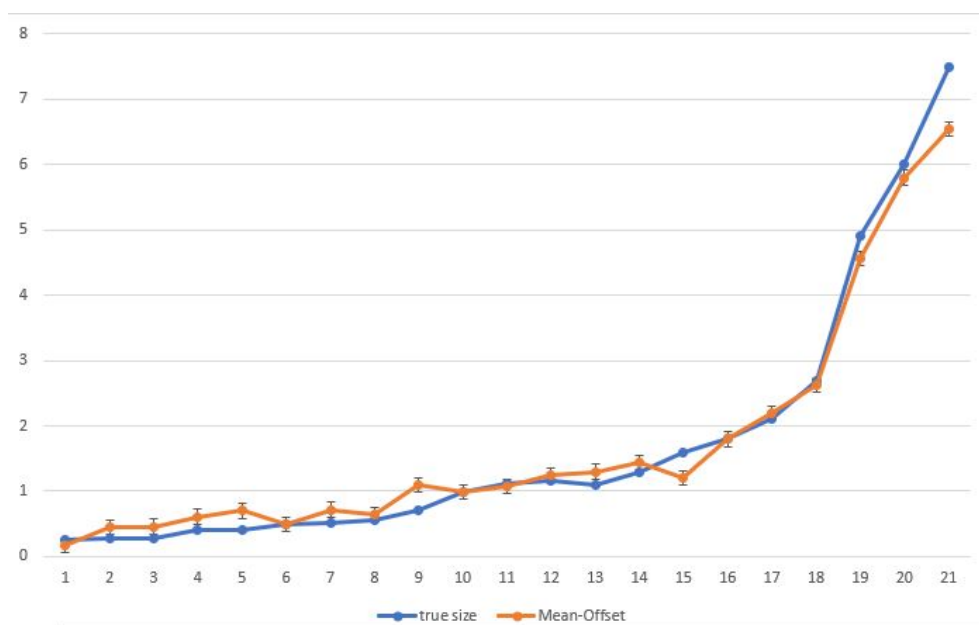


Figure 5: Plot showing the true size and mean-offset size plotted side by side

Visually we can see this is a far better fit, with far more data points now being within one standard error of the true size, however this still does not perfectly resolve the issue of the sizes the technicians achieved being so different from the true size. This leads us back to our other two hypotheses. Beginning with the technique possibly undersizing or oversizing defects at certain, unfortunately without being able to look at a greater range of defects it will not be possible to calculate or see if the technique begins to fail after a certain size. The remaining reason for the differences in the mean and the true size is the technician's ability.

In order to assess the technician's accuracy I will first assess the technician's accuracy without the offset discussed previously, and then I will look at it with the offset applied. The first step to assess the technician's accuracy is to convert their values into percentage differences from the true size. This is so the mean is not biased in favour of the larger results, allowing us to see by how much a technician over or undersizes on average. Given the risk that is incurred from a technician undersizing I hypothesise that most technicians will tend to oversize in addition to there being a positive offset due to a possible zero error. The data is available in Figure 6.

		% Difference									
A	0.25	20	100	-100	100	-100	300	100	300	220	100
B	0.27	344.4444	344.4444	85.18519	270.3704	-100	270.3704	85.18519	270.3704	196.2963	159.2593
C	0.28	150	257.1429	328.5714	78.57143	-100	257.1429	78.57143	257.1429	364.2857	185.7143
D	0.4	100	25	150	150	150	200	50	150	275	125
E	0.4	125	200	150	175	150	250	-50	150	300	150
F	0.5	0	80	100	60	100	120	-100	140	100	60
G	0.52	34.61538	188.4615	111.5385	92.30769	188.4615	73.07692	92.30769	92.30769	73.07692	92.30769
H	0.55	-27.2727	118.1818	45.45455	118.1818	63.63636	-9.09091	136.3636	81.81818	190.9091	81.81818
I	0.7	42.85714	114.2857	28.57143	71.42857	128.5714	185.7143	128.5714	42.85714	214.2857	100
J	0.98	42.85714	2.040816	-18.3673	63.26531	104.0816	93.87755	-48.9796	2.040816	83.67347	32.65306
K	1.12	16.07143	-55.3571	-10.7143	60.71429	60.71429	-10.7143	60.71429	33.92857	78.57143	25
L	1.15	4.347826	56.52174	13.04348	30.43478	-13.0435	91.30435	56.52174	30.43478	73.91304	39.13043
M	1.1	45.45455	81.81818	27.27273	36.36364	45.45455	81.81818	-54.5455	81.81818	100	45.45455
N	1.3	30.76923	61.53846	-30.7692	53.84615	53.84615	69.23077	30.76923	15.38462	53.84615	30.76923
O	1.6	-6.25	-50	-43.75	-6.25	25	25	-25	25	25	0
P	1.8	-38.8889	11.11111	-22.2222	77.77778	11.11111	33.33333	22.22222	38.88889	38.88889	16.66667
Q	2.1	23.80952	-14.2857	9.52381	42.85714	19.04762	66.66667	-4.7619	-4.7619	52.38095	19.04762
R	2.7	-18.5185	11.11111	18.51852	29.62963	-7.40741	3.703704	-7.40741	29.62963	29.62963	11.11111
S	4.9	12.2449	-8.16327	-8.16327	-10.2041	-8.16327	16.32653	-14.2857	12.2449	10.20408	0
T	6	0	-8.33333	3.333333	-8.33333	8.333333	16.66667	-11.6667	8.333333	11.66667	3.333333
U	7.5	-13.3333	-20	-16	-13.3333	-4	6.666667	-13.3333	6.666667	-8	-6.66667

Figure 6: Table showing percentage deviation from the true size

Using the data in Figure 6 I have used Excel's Descriptive Statistics function to calculate the mean and standard deviation, as shown in Figure 7 Looking at this data we can see that technician 7 has the most precise results as their mean is the closest to zero, however technician 10 is slightly more accurate as they have the smallest standard error. Technician 10 does however have a large positive mean, suggesting they do oversize results. As was predicted earlier all technicians do currently oversize defect, however this may be due to the positive offset discussed earlier. Looking at the range we see that technician 10 has the smallest range of results, suggesting that they oversize and undersize by the smallest amount, technician 7 also has a smaller range than most of the technicians, leading me to conclude that technicians 7 and 10 are the most precise and accurate. respectively.

	Op1	Op2	Op3	Op4	Op5	Op6	Op7	Op8	Op9	Op10
Mean	42.29562	71.21516	39.0965	70.12514	36.93542	101.9568	24.34508	84.00499	118.268	60.5047
Std Error	18.34161	22.78593	20.05515	14.82527	17.82336	21.99949	14.18804	20.41212	23.23959	12.62753
Range	383.3333	399.8016	428.5714	283.7037	288.4615	310.7143	236.3636	304.7619	372.2857	192.381

Figure 7: Table showing the mean and standard error of each technician.

From the data in Figure 7 we can see that technicians 1,3,5,7 and 10 have below average means for their group, meaning they are the most precise technicians. Technicians 1,4,5,7 and 10 have below average standard errors for their group, meaning they are the most accurate technicians. Technicians 4,5,6,7,8 and 10 have below average ranges for their group meaning they have the greatest consistency. From this short analysis we can see that technicians 5, 7 and 10 are the top performers as they excell in each category.

In Figure ?? I have applied the offset to account for the zero error I believe has decreased the precision of the technicians. I shall perform the same analysis done in Figure 7 and compare it with the results.

	Op1	Op2	Op3	Op4	Op5	Op6	Op7	Op8	Op9	Op10
0.25	-116.7619048	-36.7619048	-236.7619048	-36.7619048	-236.7619048	163.2380952	-36.76190476	163.2380952	83.23809524	-36.7619048
0.27	217.8130511	217.8130511	-41.44620811	143.7389771	-226.6313933	143.7389771	-41.44620811	143.7389771	69.664903	32.62786596
0.28	27.89115646	135.0340136	206.462585	-43.537415	-222.1088435	135.0340136	-43.53741497	135.0340136	242.1768707	63.60544218
0.4	14.52380952	-60.4761905	64.52380952	64.52380952	64.52380952	114.5238095	-35.47619048	64.52380952	189.5238095	39.52380952
0.4	39.52380952	114.5238095	64.52380952	89.52380952	64.52380952	164.5238095	-135.4761905	64.52380952	214.5238095	64.52380952
0.5	-68.38095238	11.61904762	31.61904762	-8.38095238	31.61904762	51.61904762	-168.3809524	71.61904762	31.61904762	-8.38095238
0.52	-31.13553114	122.7106227	45.78754579	26.55677656	122.7106227	7.326007326	26.55677656	26.55677656	7.326007326	26.55677656
0.55	-89.43722944	56.01731602	-16.70995671	56.01731602	1.471861472	-71.2554113	74.1991342	19.65367965	128.7445887	19.65367965
0.7	-5.986394558	65.44217687	-20.27210884	22.58503401	79.72789116	136.8707483	79.72789116	-5.98639456	165.4421769	51.15646259
0.98	7.968901846	-32.8474247	-53.25558795	28.37706511	69.19339164	58.98931001	-83.86783285	-32.8474247	48.78522838	-2.23517979
1.12	-14.45578231	-85.8843537	-41.2414966	30.18707483	30.18707483	-41.2414966	30.18707483	3.401360544	48.04421769	-5.52721088
1.15	-25.38302277	26.79089027	-16.6873706	0.703933747	-42.77432712	61.57349896	26.79089027	0.703933747	44.18219462	9.399585921
1.1	14.37229437	50.73593074	-3.80952381	5.281385281	14.37229437	50.73593074	-85.62770563	50.73593074	68.91774892	14.37229437
1.3	4.468864469	35.23809524	-57.06959707	27.54578755	27.54578755	42.93040293	4.468864469	-10.9157509	27.54578755	4.468864469
1.6	-27.61904762	-71.3690476	-65.11904762	-27.6190476	3.630952381	3.630952381	-46.36904762	3.630952381	-21.3690476	-21.3690476
1.8	-57.88359788	-7.88359788	-41.21693122	58.78306878	-7.883597884	14.33862434	3.227513228	19.89417989	19.89417989	-2.32804233
2.1	7.528344671	-30.5668934	-6.757369615	26.57596372	2.766439909	50.38548753	-21.0430839	-21.0430839	36.09977324	2.766439909
2.7	-31.18165785	-1.55202822	5.855379189	16.9664903	-20.07054674	-8.95943563	-20.07054674	16.9664903	-1.55202822	-1.55202822
4.9	5.267249757	-15.1409135	-15.14091351	-17.1817298	-15.14091351	9.34888241	-21.26336249	5.267249757	3.226433431	-6.9776482
6	-5.698412698	-14.031746	-2.365079365	-14.031746	2.634920635	10.96825397	-17.36507937	2.634920635	5.968253968	-2.36507937
7.5	-17.89206349	-24.5587302	-20.55873016	-17.8920635	-8.558730159	2.107936508	-17.89206349	2.107936508	-12.55873016	-11.2253968
	Op1	Op2	Op3	Op4	Op5	Op6	Op7	Op8	Op9	Op10
Mean	-7.259910244	21.65962492	-10.45903092	20.56960157	-12.62011208	52.40130688	-25.21044945	34.44945282	68.71246851	10.94916858
Std Error	13.39268685	15.89148816	16.44750085	9.322531034	20.46208302	14.03688575	12.50493321	11.41181733	15.60981975	5.592810982
Range	334.5749559	303.6974049	443.2244898	187.276392	359.4725275	235.7792208	248.1088435	196.0855199	254.7356009	101.2857143

Figure 8: Updated table with offset applied.

In Figure 8 we can see that technicians 1,3,5 and 7 have negative means with only technician 7 being beyond their standard error to the 0 mark and thus is likely undersizing defects. Another effect of applying the offset is that it has now decreased the standard error of all technicians has decreased, meaning all technicians are more accurate when the offset is applied. Almost all of the technicians have had a decreased range, the only exception being technician 3 whose range has increased from 428.5714 to 443.2244, this means most technicians are more consistent when the theoretical zero error is accounted for. Technicians 1,3,5 and 10 have got the means closest to zero, this means they are the most precise technicians. Technicians 1,4,6,7,8 and 10 have the smallest standard errors and thus are the most precise technicians. Technicians 4,6,7,8,9,10 have the smallest ranges and are thus the most consistent technicians. Technician 10 is now the only technician who excels in all categories is likely the most accurate technician.

5 Conclusions

From my analysis I have found that it is highly likely there is a zero error in the data causing an positive offset of 0.34mm. When the offset is accounted for, as shown in Figure 5, the mean defect sizing performed by the technicians aligns better with the true defect size. It is not possible to discern whether the sizing technique is affected by the size of the defect as the dataset is too small, and fits too well when the offset is applied. It is however possible that the sizing technique could be affected if we continued sizing defect beneath the minimum and maximum values and would be an interesting investigation. Without considering the zero error the most accurate technicians were 5,7 and 10, with 7 being the most precise and 10 the most accurate. Accounting for the zero error we find that only technician 10 excels in categories and thus is the best technician. Without accounting for the zero error the least precise technician was technician 6 and the least accurate was technician 9. Accounting for the zero error I found the least precise technician was still technician 9 and least precise was technician 5.

As a whole the new ultrasonic gauging procedure is fairly accurate and correlates well to the actual data when the zero error is accounted for however the data provided does not allow the generation of a probability of detection graph as there is no opportunities for rejection and there is only one sample per defect size. This would benefit the method as it would allow people attempting the new technique to know what defect sizes they are likely to detect.

References

- [1] Keprate, A. (2016). Probability of Detection: History, Development and Future. [online] Available at: https://www.researchgate.net/publication/299603630.Probability_of_Detection_History_Development_and_Future [Accessed 18 Nov. 2021].
- [2] NONDESTRUCTIVE EVALUATION REQUIREMENTS FOR FRACTURE-CRITICAL METALLIC COMPONENTS. (2008). [online] Washington, DC: National Aeronautics and Space Administration. Available at: https://standards.nasa.gov/sites/default/files/nasa_std_5009.pdf [Accessed 2 Jan. 2022].
- [3] Probability of Detection (PoD) curves Derivation, applications and limitations. (2006). [online] Jacobi Consulting Limited. Available at: <https://www.hse.gov.uk/research/rrpdf/rr454.pdf> [Accessed Jan. 1AD].
- [4] Vrana, J., Zimmer, A., Bailey, K. and Angal, R. ... (2010). Evolution of the Ultrasonic Inspection Requirements of Heavy Rotor Forgings Over the Past Decades.

- [5] Vrana, J., Zimmer, A., Lohmann, H.-P. and Heinrich, W. (2016). Evolution of the Ultrasonic Inspection over the Past Decades on the Example of Heavy Rotor Forgings. [online] Available at: <https://www.ndt.net/article/wcndt2016/papers/th2a5.pdf> [Accessed Jan. 1AD].
- [6] Ditchburn, R. and Ibrahim, M. (2009). Ultrasonic Phased Arrays for the Inspection of Thick-Section Welds.
- [7]