ARL-TR-9326 ● Sep 2021

Comparing Thickness Measurement Methods Using Bland–Altman Analysis

by Lucas Tommervik, Daniel Shreiber, and Daniel Pope

NOTICES

Disclaimers

The findings in this report are not to be construed as an official Department of the Army position unless so designated by other authorized documents.

Citation of manufacturer's or trade names does not constitute an official endorsement or approval of the use thereof.

Destroy this report when it is no longer needed. Do not return it to the originator.

ARL-TR-9326 ● Sep 2021

Comparing Thickness Measurement Methods Using Bland–Altman Analysis

Lucas Tommervik
Oak Ridge Associated Universities

Daniel Shreiber and Daniel Pope
Weapons and Materials Research Directorate,
DEVCOM Army Research Laboratory

REPORT DOCUMENTATION PAGE	*Form Approved* *OMB No. 0704-0188*

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing the burden, to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.

PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.

1. REPORT DATE *(DD-MM-YYYY)* September 2021	**2. REPORT TYPE** Technical Report	**3. DATES COVERED (From - To)** August 2020–December 2020

4. TITLE AND SUBTITLE Comparing Thickness Measurement Methods Using Bland–Altman Analysis	**5a. CONTRACT NUMBER** W911NF-20-2-0033
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S) Lucas Tommervik, Daniel Shreiber, and Daniel Pope	**5d. PROJECT NUMBER**
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) DEVCOM Army Research Laboratory ATTN: FCDD-RLW-ME Aberdeen Proving Ground, MD 21005	**8. PERFORMING ORGANIZATION REPORT NUMBER** ARL-TR-9326
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	**10. SPONSOR/MONITOR'S ACRONYM(S)**
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release: distribution unlimited.

13. SUPPLEMENTARY NOTES

ORCID ID: Daniel Pope, 0000-0001-9133-7810

14. ABSTRACT

This report discusses the Bland–Altman method of statistical comparison and employs it to compare and contrast two different methods of chemical agent resistant coating (CARC) thickness measurement, implementing this method of statistical comparison for the first time outside the medical sciences context. The same CARC samples were separately measured using a Traceable digital caliper and an Elcometer 456 coating thickness gauge. Measuring CARC thickness accurately is of the utmost importance to accurately determine the dielectric properties of a given CARC sample and consequently use the dielectric responses of CARCs to detect vulnerabilities, weathering, and any inhomogeneity in the deposited coating. Using the Bland–Altman method of statistical comparison demonstrates the digital calipers are not interchangeable with the Elcometer when it comes to measuring the thickness of CARC coatings and shows the digital calipers consistently overestimate the thickness by 5.94 µm. The viability and effectiveness of the Bland–Altman method of statistical comparison outside a medical sciences context are also demonstrated and affirmed.

15. SUBJECT TERMS

Bland–Altman method of statistical comparison, CARCs, thickness measurements, dielectric properties, terahertz time-domain spectroscopy

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON** Daniel Shreiber
a. REPORT Unclassified	**b. ABSTRACT** Unclassified	**c. THIS PAGE** Unclassified	UU	23	**19b. TELEPHONE NUMBER (Include area code)** 410-306-4928

Contents

List of Figures

Acknowledgments

The author would like to offer thanks to Oak Ridge Associated Universities for providing the funding necessary for this work as well as all other unpublished, forthcoming works.

Lucas Tommervik was sponsored by the DEVCOM Army Research Laboratory under Cooperative Agreement No. W911NF-20-2-0033. The view and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of DEVCOM ARL or the US Government. The US Government is authorized to reproduce and distribute reprints for Government purposes not withstanding any copyright notation herein.

1. Introduction

Chemical agent resistant coatings (CARCs) are included on all deployable tactical equipment used by the US military and are important to their operation. CARC maintenance is of utmost importance to keep tactical equipment field ready. Currently, there are no adequate, nondestructive methods of evaluation when it comes to the performance of deposited CARCs. Being able to determine whether a CARC deposited on a military platform will perform as expected or whether certain defects, such as inhomogeneous mixing or substrate interface corrosion, are present is of the utmost importance. Unfortunately, many current methods of evaluation of CARC performance are destructive in nature. That is, testing the viability of a CARC usually involves damaging the CARC in the process or inducing corrosion on the underlying substrate to evaluate its performance. A nondestructive method of evaluation that would allow CARC researchers to correlate identified defects in a CARC with its performance while preserving the integrity of the coating is crucial. Terahertz time-domain spectroscopy (THz-TDS) invites a new nondestructive method of evaluation of CARC viability. In particular, spectroscopic analysis of a THz pulse transmitted through a material unlocks the potential of identifying and characterizing that material's parameters, such as the dielectric properties.[1,2] This extraction of dielectric properties, in turn, allows for the creation of a "library" of accepted dielectric characteristics for any given CARC. Eventually, this library would become the reference against which to evaluate CARCs in the field. That is, if the dielectric properties of a CARC in the field deviate from the standard, degradation of the CARC by one means or another can be assumed and that piece of equipment can be tagged for maintenance. Likewise, local deviations in a singular CARC's dielectric properties would indicate a local inhomogeneity in the deposited coating, allowing for targeted maintenance to occur.

In creating this library, we must determine the complex refractive index of a multitude of CARC samples. Part of this analytical process requires a determination of the thickness of a given CARC sample. The accuracy of our thickness measurement is directly related to the accuracy of our refractive index measurement. The refractive index is not dependent on the thickness measurement. That is, the dielectric properties of the CARC will always be the same in theory. Thus, how thick the sample is should not influence the outcome of the analysis of its dielectric properties. For example, a 90-µm sample and a 210-µm sample of the same CARC should yield the same refractive index. However, this does not diminish the importance of obtaining an accurate thickness, as it is imperative to determining the correct refractive index of the CARC. A bad thickness measurement, independent of what the thickness value is, will still yield an incorrect refractive index value. Accordingly, it is extremely important that we choose the correct method of measuring the thickness of the CARC samples.

Two methods of thickness measurement were examined: a Traceable digital caliper and an Elcometer 456 coating thickness gauge. The Elcometer relies on eddy currents to make its thickness measurements. To discern which method of thickness measurement is best, the Bland–Altman method of statistical comparison between two different methods of measurement was used. This method of statistical comparison was actually first employed in the medical sciences, but because it is solely reliant on statistical models, it is equally applicable to any analytical science. The Bland–Altman method is designed to compare a new, untested measurement method to an established, accepted measurement method by, in short, comparing the differences in measured values between the two methods when used on the same sample. The analysis does not tell you which method to use of the two, but rather, simply provides an answer on whether the new, untested method of measurement is interchangeable with the established method.[3] If interchangeable, the user is left with a choice of which method to use. If not interchangeable, then the established method is deemed superior to the new method. Despite not having a new and established method in this case, the Elcometer has been dubbed as the "established" method of measurement and the digital caliper as the "new" method of measurement. This is, in part, because the Elcometer is specifically designed to operate as a coating thickness measurement method by its manufacturer. In addition, the Elcometer is the more accurate of the two methods of measurement (±2.5 µm vs. ±30 µm). As such, we have endeavored to prove whether the digital caliper method of measuring thickness is interchangeable with the Elcometer using the Bland–Altman method of statistical comparison.

2. Experimental Layout

As stated previously, the two methods of thickness measurement examined were a Traceable digital caliper[4] with an accuracy of ±30 µm and an Elcometer 456 coating thickness gauge[5] with an accuracy of ±2.5 µm. The experimental setup was simple. First, a single point was marked on each CARC sample as the point of measurement. If different points on the sample had been measured, the thickness measurements could have fallen victim to the thickness variations in a single coating. Next, at this marked point, the thickness of the coating was simply measured with both the digital caliper and the Elcometer. For the digital caliper, the coating sample was clasped in the caliper to obtain a thickness measurement. The Elcometer requires a reference with which to compare the thickness of the CARC. Accordingly, in measuring the thickness of the samples with the Elcometer, the samples were required to be placed on a piece of bare Aluminum 2024. The Elcometer device was then calibrated to treat this substrate as a reference before making the thickness measurements. Finally, the Menlo TeraK15 terahertz spectrometer and its associated software were used to perform spectroscopic analysis at the predetermined point on the sample and obtain the refractive index of the given CARC.

3. Why Not Other Methods?

The bulk of the work in this report involved performing a statistical analysis upon the thickness measurement methods. One must understand what the Bland–Altman method is truly measuring before delving into the methodology. A natural tendency when statistically comparing two methods of measurement is to plot them against one another, perform linear regression, and calculate their correlation coefficient to determine how the two methods are related. This can be seen for the thickness measurement data in Fig. 1. However, the correlation coefficient and corresponding linear regression can be a misleading statistic. The correlation coefficient, r, only displays how the two variables measured are related as a value from 0 to 1. It does not, however, display the level of agreement between the two variables.[3,6] Assuming the correlation is statistically significant to the 95% significance level, $p < 0.05$, the most the correlation coefficient can tell us is the level to which the two variables are related. The correlation coefficient cannot, however, tell us whether the relation is equivalent to agreement, except in one specific case. So, given these constraints, the correlation coefficient can really only give us one of three outcomes: perfect agreement, no relation, or a more nebulous relation.

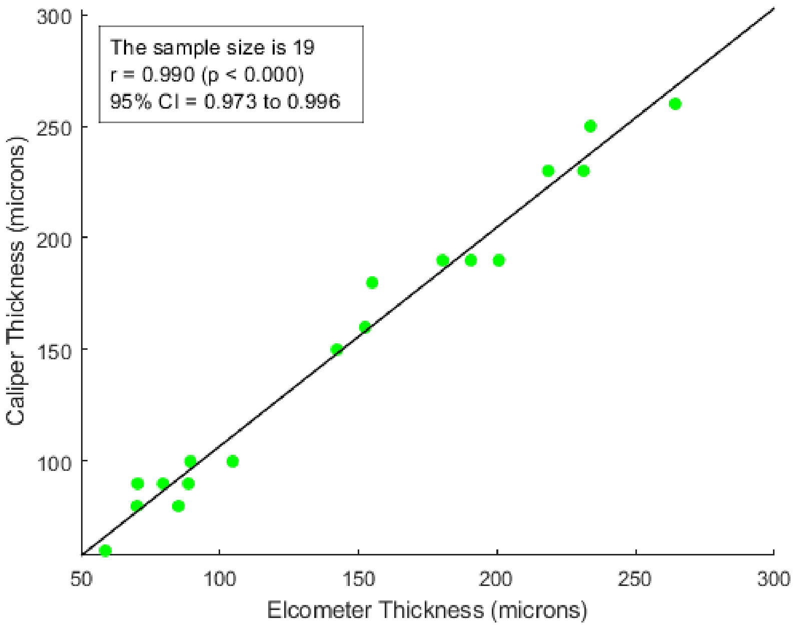

Fig. 1 Linear regression comparing Elcometer thickness measurements to caliper thickness measurements including the correlation coefficient and significance level

Examining these three outcomes under the lens of the experiment, the first outcome would be perfect agreement: r = 1, both the Elcometer and digital caliper measure the exact same thickness for every single sample, and every point in Fig. 1 would lie along the linear regression line.[3] The measured r is r = 0.990, meaning there is not perfect agreement between the Elcometer and the digital caliper. This, of course, makes sense because the accuracy of the two devices will introduce natural error on their own so there should not be perfect agreement.[6] The second possible outcome would be no relation: the significance level, p, would indicate statistical significance but the correlation coefficient would be closer to r = 0. In this case, it could be claimed there is no significant relationship between the thickness measurements of the Elcometer and the digital caliper and there would be separate tests conducted to see which method provides the best thickness measurements. The measured significance level is p = $7.0394e^{-16}$ (rounded to 0.000 in Fig. 1) indicating a statistically significant correlation coefficient. However, again, r = 0.990, which is extremely close to 1 and would not allow for a claim of a lack of a significant relationship. This leaves the third option: r does not equal 1 and p lends itself to statistical significance. This is the situation this work deals with. However, all this allows to be claimed is that there is a statistically significant relationship between the Elcometer's thickness measurements and the digital caliper's thickness measurements. Since both methods of measurement are measuring thickness it would of course make sense that the two methods are related to one another, but this correlation test does little to show the agreement between the two methods.[3]

4. The Bland–Altman Method of Statistical Comparison: Explanation through Example

As a result of the correlation test only illuminating the relationship between two separate methods of measurement, Bland and Altman developed a statistical model that would aid in the understanding of the agreement between two different measurement methods. Bland–Altman analysis is more easily explained through an example, so the method will be explained while discussing the results obtained in this work. The essence of a Bland–Altman plot is the difference between the new method of measurement and the established method of measurement. The simplest way to quantify how two different methods of measurement agree or differ is, quite literally, the difference between them. As can be seen in Fig. 2, the x-axis of a Bland–Altman scatter plot is the average measurement of each sample based on the two methods of measurement while the y-axis is the difference between the new method of measurement and the established method of measurement.

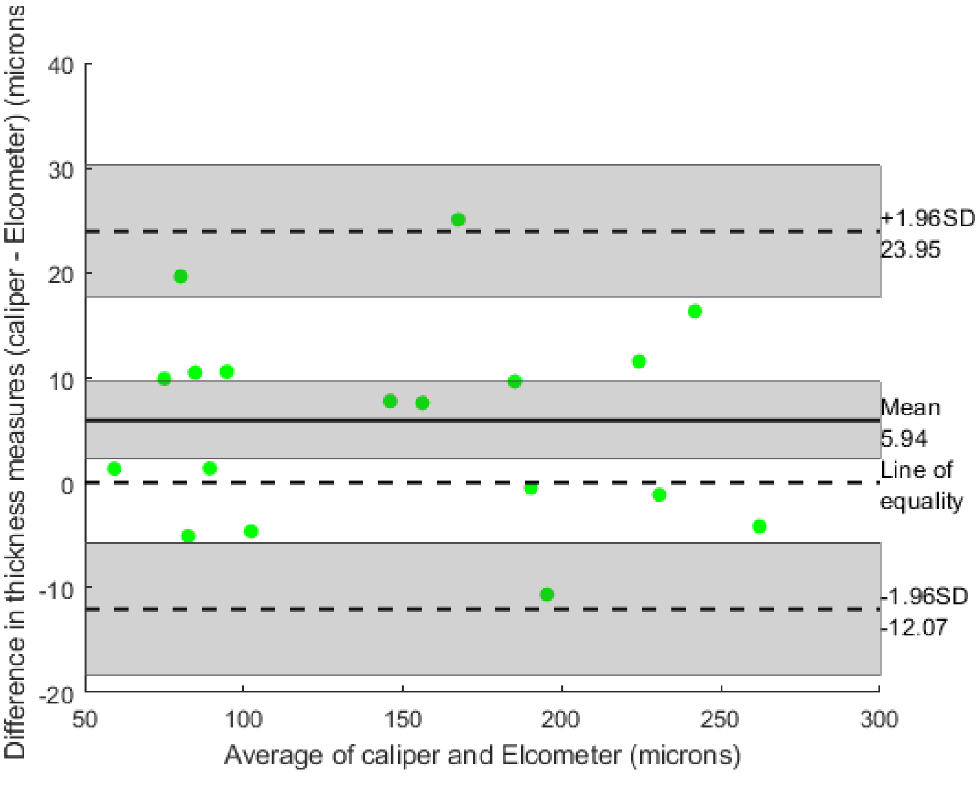

Fig. 2 Bland–Altman plot for comparison of digital caliper's thickness measurements and Elcometer's thickness measurements. The solid line displays the bias. The upper and lower dashed lines display the limits of agreement while the middle dashed line is the line of equality. The gray boxes represent the 95% confidence intervals for their respective variables.

The two most important values gleaned from a Bland–Altman analysis are the bias and the limits of agreement. As it was claimed when looking at the correlation coefficient and linear regression line, if there had been perfect agreement between the two methods, then every single measurement of the same sample would have yielded the same thickness regardless of the method used. In other words, the average difference between the caliper and the Elcometer when measuring thickness, the bias, would have been zero. This fact is visualized in Fig. 2 by the line denoted line of equality.[6] The bias, in essence, gives a way to quantify exactly how the two methods of measurement agree with one another. The bias shows us, on average, how much the new method of measurement differs from the established method of measurement. So, for instance, in Fig. 2 it is shown that on average the digital caliper is measuring a thickness of a given CARC 5.94 µm greater than the thickness measured by the Elcometer. This means the calipers have a positive bias when compared with the Elcometer—they tend to measure a higher value for any given sample.[6,7] Likewise, a negative bias is possible. A negative bias implies that, on average, the new method of measurement measures a value less than the established method of measurement—the new method tends to measure a lower value for any given sample.

The limits of agreement also give us information about the agreement between the two methods of measurement. More specifically, they tell us about the precision and reliability of the new method of measurement.[7] The limits of agreement take advantage of the standard deviation of the differences between the two measurement methods. In essence, assuming that the differences are normally distributed, we can claim that 95% of the difference values will lie between the limits of agreement.[3,6,7] In other words, the limits of agreement are just the 95% confidence interval for the difference values. Normal distribution of the difference values can be confirmed by checking the shape and tails of a histogram, as in Fig. 3, and by conducting a test of normality such as the Shapiro–Wilk test.[6,7]

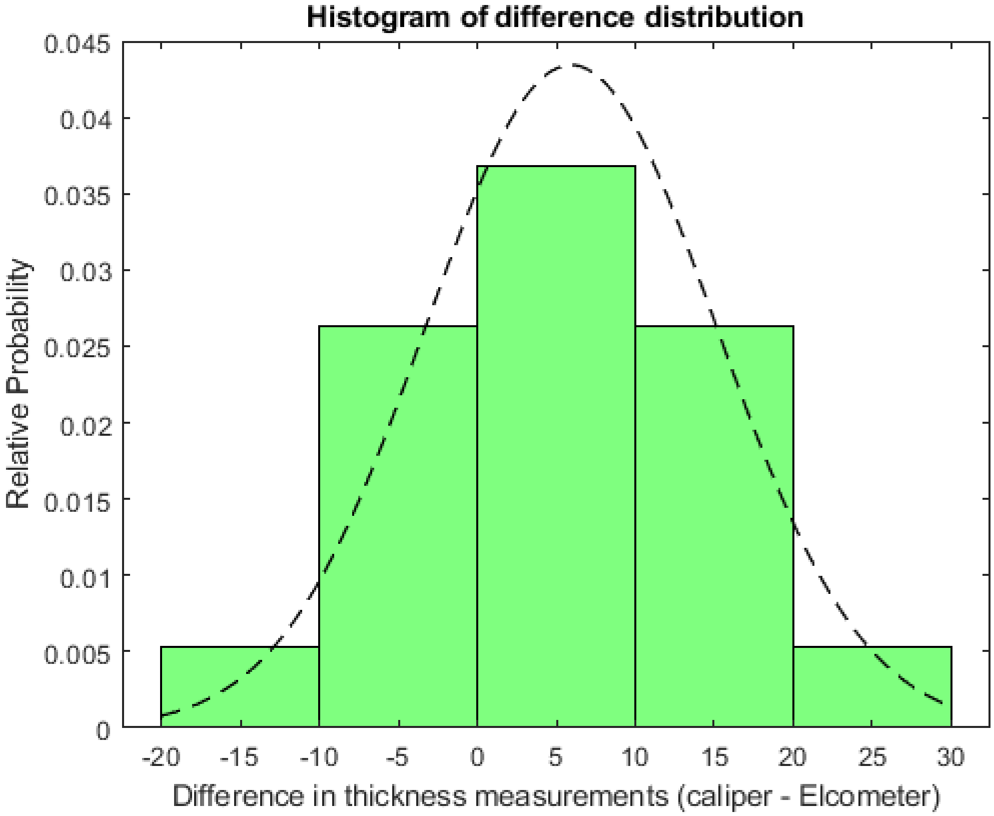

Fig. 3 Histogram used, in part, to confirm the normality of the distribution of differences between the caliper measurements and the Elcometer measurements

Now, assuming the differences between the two methods of measurement are normally distributed, we can construct the upper and lower limits of agreement using the following equation[3,6,7]:

$$(mean\ difference) \pm 1.96(standard\ deviation\ difference) \qquad (1)$$

In our case, as seen in Fig. 2, the upper limit of agreement is 23.95 µm while the lower limit of agreement is –12.07 µm. What, exactly, does this mean? It means 95% of all thickness measurements made by the digital calipers will be up to 23.95 µm more than what the Elcometer measured or 12.07 µm less than what the Elcometer measured. That is, 95% of the differences as a result of the bias in using the calipers to measure thickness are likely to lie in the range defined by the limits of agreement.[7] The limits of agreement demonstrate the precision of the new method of measurement. They show whether or not the difference between the new method of measurement, the calipers, and the established method of measurement, the Elcometer, are tightly constrained around the bias or not. This indicates whether most difference measurements are about the same or not.

One extremely important note is that the Bland–Altman analysis, for the most part, cannot and does not give you an explicit answer about the interchangeability of the two methods of measurement. The bias and limits of agreement simply quantify the accuracy (bias) and precision (limits of agreement) of the new method of measurement compared to the established method. For this reason, it is recommended to establish a priori how accurate and how precise the new method of measurement should be in order to meet the accepted criteria for interchangeability.[6,7] In the literature, how stringent the a priori criterion is entirely depends on the task at hand and how imperative it is that the correct measurement is made.[3,6,7] The investigator would then compare the calculated bias and limits of agreement with the a priori criterion to determine the viability of the new method of measurement.

The final statistics available in a Bland–Altman analysis are the 95% confidence intervals for the bias, the upper limit of agreement, and the lower limit of agreement. Simply put, these confidence intervals show the precision of the bias and limit of agreement estimates. The confidence intervals are shown in Fig. 2 as gray boxes surrounding their respective line. More specifically, the 95% confidence interval for the bias describes whether or not there is a systematic bias when measuring with the new method of measurement.[6] If the line of equality is not included in the 95% confidence interval for the bias, then we can safely claim the new method of measurement consistently measures a higher or lower value, depending on the bias.[6] For example, the line of equality is not included in the 95% confidence interval of the bias between the caliper measurements and Elcometer measurements, as seen in Fig. 2, and we can safely claim there is a systematic positive bias and the calipers will consistently overestimate the thickness of a CARC when compared with the Elcometer. The 95% confidence intervals for the limits of agreement allow us to discern the magnitude of the sampling error from the given batch of samples. That is, the wider the confidence intervals for the limits of agreement the more likely that some kind error in the sample was present.[6]

5. Results

Using the Elcometer 456 coating thickness gauge as the established method and the Traceable digital caliper as the new method, it was found that the bias of the digital caliper method of thickness measurement was 5.94 µm with a 95% confidence interval from 2.28 to 9.59 µm, as seen in Fig. 2. The bias had 95% limits of agreement from –12.07 to 23.95 µm. These limits of agreement had 95% confidence intervals of –18.40 to –5.74 µm and 17.62 to 30.28 µm, respectively. Moreover, the Shapiro–Wilk test of normality applied to the distribution of differences between the two methods yielded a p-value of 0.796, affording the claim that the differences are normally distributed and the limits of agreement for the bias are statistically significant. The sample size was 19.

Having a varied set of thickness measurements, as was done here, is important when it comes to being able to claim the results are applicable across all different types of samples.[7] Nonetheless, the samples here are also split up into “thick” and “thin” coatings so as to potentially discern whether one subset of samples was influencing the results as a whole.

When looking at the thin samples (Elcometer-measured thickness less than 105 µm), it was found that the bias of the digital caliper method of thickness measurement was 5.45 µm with a 95% confidence interval of –0.33 to 11.22 µm, as seen in Fig. 4. The bias had 95% limits of agreement from –11.45 to 22.34 µm. These limits of agreement had 95% confidence intervals of –21.45 to –1.45 µm and 12.34 to 32.34 µm, respectively. Furthermore, the Shapiro–Wilk test of normality applied to the distribution of differences between the two methods yielded a p-value of 0.431, affording the claim that the differences are normally distributed and the limits of agreement for the bias are statistically significant. The sample size for the thin samples was 8.

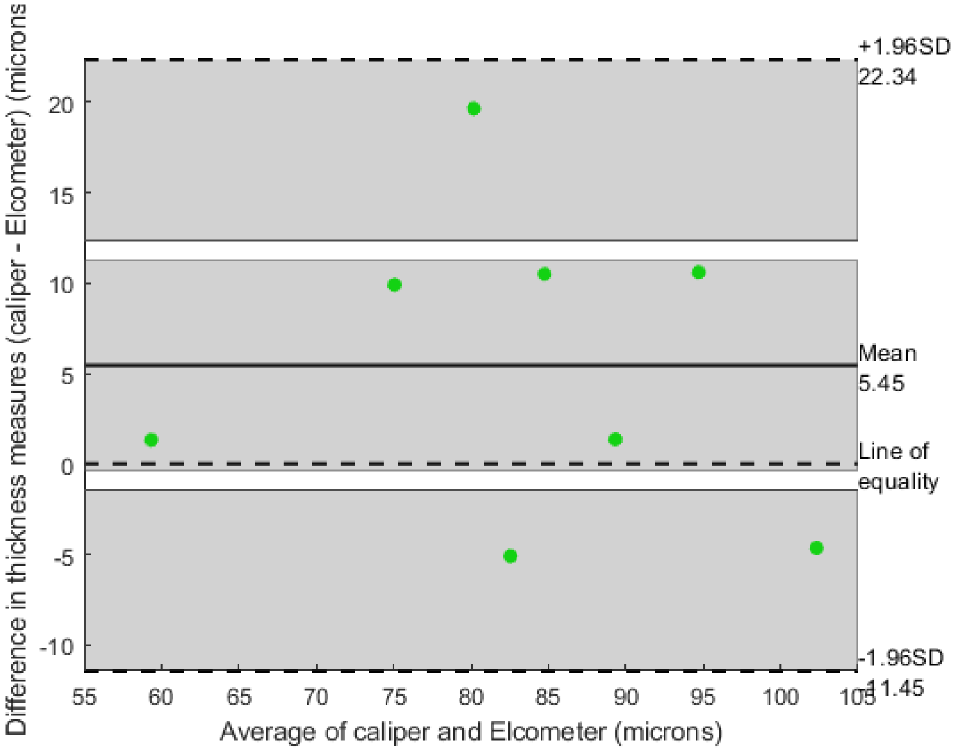

Fig. 4 Bland–Altman plot for thin samples (Elcometer-measured thickness less than 105 µm)

When looking at the thick samples (Elcometer-measured thickness greater than 105 µm), it was found that the bias of the digital caliper method of thickness measurement was 6.30 µm with a 95% confidence interval of 0.84 to 11.75 µm, as seen in Fig. 5. The bias had 95% limits of agreement from –13.26 to 25.86 µm. These limits of agreement had 95% confidence intervals of –22.71 to –3.82 µm and 16.41 to 35.30 µm, respectively. Moreover, the Shapiro–Wilk test of normality applied to the distribution of differences between the two methods yielded a p-value of 0.916, yielding a claim that the differences are normally distributed and the limits of agreement for the bias are statistically significant. The sample size for the thick samples was 11.

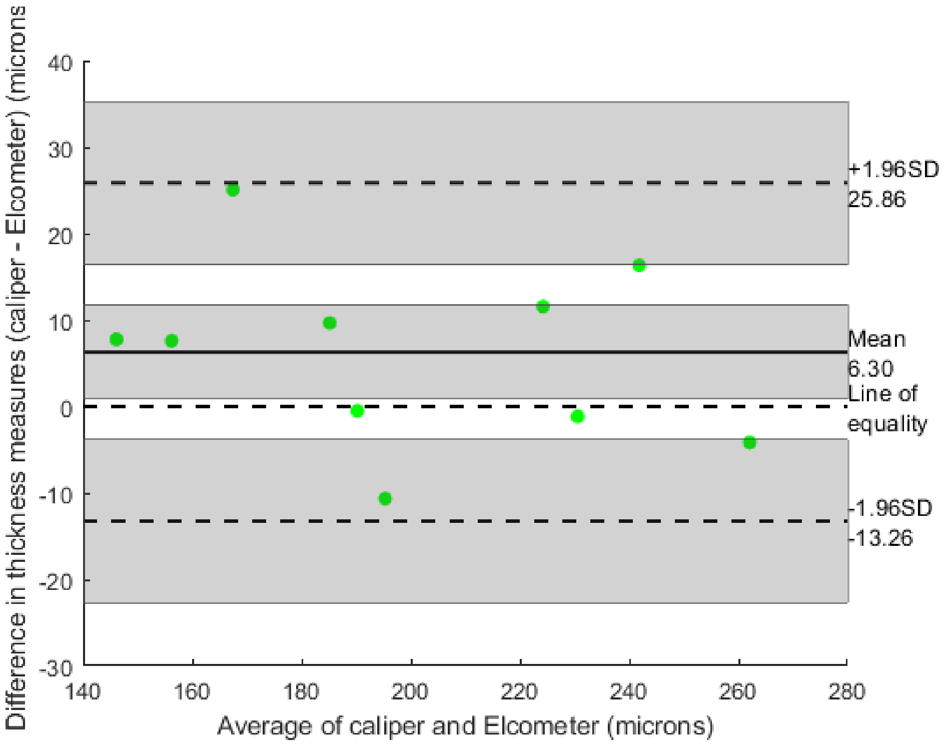

Fig. 5 Bland–Altman plot for thick samples (Elcometer-measured thickness greater than 105 µm)

6. Discussion

Before discussing the results, an a priori criterion for the accuracy and precision of our desired thickness measurement needs to be determined. It has already been determined that the accuracy of our established method, the Elcometer, is ±2.5 µm, whereas the accuracy of our new method, the digital caliper, is ±30 µm. Hence, it is reasonable to set the a priori accuracy and precision at ±30 µm. Setting the a priori criterion to be less than 30 µm would bias the results against agreement between methods due to the natural error in thickness measurements as a result of the instrument itself.[7] However, the more accurate and precise the digital calipers are, the more likely they are to be interchangeable with the Elcometer. Thus, we set the accuracy and precision to be as stringent as possible.

Now, in analyzing both the thin and thick sample subsets it can be seen that accuracy is not an issue. The bias, or accuracy, of both the thick subset at 6.30 µm and the thin subset at 5.45 µm are well under the predetermined accuracy threshold. This means that, on average, the differences in thickness measurements when using the caliper are negligible enough to overcome the inherent error present in using the digital caliper. Likewise, the precision for both the thick and thin subsets both fall underneath the acceptable precision threshold. For the thick subset it is shown that the upper limit of agreement of 25.86 µm and the lower limit of agreement of –13.26 µm both fall under the ±30 µm precision threshold. This means that, on average, the differences in thickness measurements when using the caliper are small enough to overcome any inherent error present. For the thin subset, similar results were obtained—the upper limit of agreement of 22.34 µm and the lower limit of agreement of 11.45 µm both fall under the ±30 µm threshold.

However, the issues presented by these subsets are twofold. The first issue applies only when considering the thick samples. Figure 5 shows the line of equality does not lie within the 95% confidence interval for the bias. That is, when working with thick samples, the digital calipers consistently overestimate the thickness of the CARC sample.[6] Ninety-five percent of the average differences between the two methods for thick samples are above the line on which the two methods would measure identical values. This means the digital calipers are always positively biased for thick samples.

The second issue is that both the thin and thick samples have sampling error issues. The range of the 95% confidence interval on the bias is about 10 µm for both sample subsets as seen in Figs. 4 and 5. Likewise, the range on the 95% confidence intervals for the limits of agreement for both is about 20 µm. This makes sense because when breaking the sample set into subsets, the number of samples is decreased by almost half, increasing the sampling error. However, this is an issue because the true bias and true limits of agreement could be much higher or lower than calculated. In particular, the confidence interval on the upper limit of agreement for both the thick and thin subsets even extends past our threshold for precision. The sampling error is too great in this case to make any specific claims about the agreement between the measurement methods using only the sample subsets.

As a result, it is better to analyze the entire sample set. As mentioned previously, the entire sample set is also a greater indicator of agreement between methods because varied sample sets allow us to make claims about the two different methods of measurement for typical, varied sample sets. For the entire sample it has been shown that the bias is 5.94 µm, the upper limit of agreement is 23.95 µm, and the lower limit of agreement is –12.07 µm. The bias falls underneath the accuracy threshold and the limits of agreement fall underneath the precision threshold, indicating good agreement between the digital caliper method of thickness measurement and the Elcometer method of thickness measurement. However, much like the thick subset, the line of equality does not fall within the 95% confidence interval for the bias. In fact, the lower limit of the bias confidence interval is farther from the line of equality than it is for the thick subset. This indicates that, when viewing the entire sample set, the positive systematic bias is even more apparent. The digital calipers consistently overestimate the thickness of a given CARC when compared with the Elcometer.

Furthermore, when viewing the entire sample set, the sampling error, while still present, is much less egregious. The upper limit of the confidence interval for the upper limit of agreement only passes the threshold for precision in the 95th percentile. As a result, the sampling error is not too great in this case to make any claims about agreement. In fact, this statistical comparison could potentially indicate the digital calipers tend to be accurate and precise enough to be used interchangeably with the Elcometer. However, the extremely crucial fact that the line of equality and the confidence interval of the bias do not overlap determines that, to 95% confidence, the digital calipers consistently overestimate the thickness value and lend themselves to a positive systematic bias. Therefore, the presented research indicates the digital calipers are not to be used interchangeably with the Elcometer.

7. Conclusion

The presented research demonstrates that the digital calipers are not interchangeable with the Elcometer due to positive systematic bias causing consistent overestimation of the thickness values when using the calipers. Now that the Elcometer has been established as the best thickness measurement method, it will be possible to continue to confidently and accurately construct the library of CARC dielectric properties. In the future, when completing scans of coatings applied to substrates, the correct dielectric values will be known a priori. As a result, by testing for dielectric properties with THz-TDS, the capability to discern any chemical inhomogeneity when measured dielectric values differ from the library is enhanced. Furthermore, this report establishes the Elcometer as the method of coating thickness measurement. As a result, this same statistical methodology can be applied to compare the Elcometer against any future thickness measurement methods. Finally, it has been shown that the Bland–Altman method of statistical comparison is a viable and effective method outside the field of medical sciences and can be applied to a wide variety of fields in the future.

8. References

1. Dorney TD, Baraniuk RG, Mittleman DM. Material parameter estimation with terahertz time-domain spectroscopy. Journal of the Optical Society of America A. 2001;18(7):1562.

2. Duvillaret L, Garet F, Coutaz J-L. A reliable method for extraction of material parameters in terahertz time-domain spectroscopy. IEEE Journal of Selected Topics in Quantum Electronics. 1996;2(3):739–746.

3. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. The Lancet. 1986;327(8476):307–310.

4. Fisherbrand Traceable Digital Calipers. Fisher Scientific [accessed 2021 Sep 30]. https://www.fishersci.ca/shop/products/fisherbrand-traceable-digital-calipers-2/p-166031.

5. Elcometer 456 Coating Thickness Gauge. Elcometer [accessed 2021 Sep 30]. https://www.elcometer.com/en/coating-inspection/dry-film-thickness/dry-film-thickness-digital/elcometer-456-coating-thickness-gauge.html.

6. Giavarina D. Understanding Bland Altman analysis. Biochemia Medica. 2015;25(2):141–151.

7. Hanneman SK. Design, analysis, and interpretation of method-comparison studies. AACN Advanced Critical Care. 2008;19(2):223–234.

List of Symbols, Abbreviations, and Acronyms

ARL	Army Research Laboratory
CARC	Chemical Agent Resistant Coating
DEVCOM	US Army Combat Capabilities Development Command
TDS	time-domain spectroscopy
THz	terahertz

1 DEFENSE TECHNICAL
(PDF) INFORMATION CTR
DTIC OCA

1 DEVCOM ARL
(PDF) FCDD RLD DCI
TECH LIB

3 DEVCOM ARL
(PDF) FCDD RLW ME
L TOMMERVIK
D SCHREIBER
FCDD RLW MC
D POPE