Statistical Analysis of Impulsive Flashover Voltages Across Solid-Air Interfaces

R.W. Macpherson, M. P. Wilson, M. J. Given, I.V. Timoshkin, S. J. MacGregor
Dept. Electronic & Electrical Engineering, University of Strathclyde, 204 George Street,
Glasgow, G1 1XW, UK

***Abstract*- Understanding of the influence of experimental conditions on the breakdown voltage of composite insulation is important to facilitate optimal design of high voltage systems. This includes ensuring that the statistical analysis performed on breakdown voltage data is relevant, in providing extra information on the failure voltages. Therefore, in order to investigate the applicability of statistical techniques in aiding to elucidate further detail about the breakdown process, multiple statistical methods were applied and analysed, in order to find the most suitable for the data found during a specific set of breakdown tests. Normal, lognormal, 2-parameter Weibull and 3-parameter Weibull fittings are discussed herein, as applied to the authors' experimental data on the flashover voltages across solid-air interfaces, subjected to 100/700 ns impulse voltages. Negative impulse voltages were applied to composite insulation, consisting of one of three different solid materials in air - Delrin (Polyoxymethylene), Ultem (Polyetherimide) and HDPE (High Density Polyethylene). Samples of each material were machined to a smooth finish. The environmental conditions used in this paper were a fixed air pressure of −0.5 bar gauge, and relative humidity (RH) levels of <10% RH, ~50% RH and >90% RH. The data points used in the statistical analyses were from the average flashover voltages found using the ASTM D3426-97 'step up' testing procedure. Fitting the normal, log-normal, 2-parameter Weibull and 3-paramter Weibull distributions to breakdown voltage data obtained for each set of test conditions allowed for the relative quality of fit of each to be directly compared. The Kolmogorov-Smirnov (K-S) test was deployed in order to compare the maximum distance between the experimental data and the theoretical cumulative distribution function, thus verifying the most accurate method of statistical analysis for a given dataset. It was found that the overall best fit to the experimental data was given by the 3-parameter Weibull distribution.**

I. INTRODUCTION

Within the pulsed power industry, a key factor determining the achievable output voltage of a HV system is the flashover voltage of the insulating parts. Statistical analysis of the breakdown voltages associated with solid-gas interfaces can reveal useful information to aid system designers in the selection of solid materials. However, it is important to test the applicability of the distribution being applied, to ensure that the fitting parameters obtained are truly representative of the distribution of the data. In this paper, different statistical analyses will be conducted, in order to find the distribution of best fit for a specific dataset of breakdown values achieved during an experimental testing process.

Normal, lognormal, 2-parameter Weibull and 3-parameter Weibull cumulative distribution functions (CDF) were plotted, to enable extraction of the specific fitting parameters associated with each distribution. The CDF for each statistical method has been plotted alongside the empirical cumulative distribution function (ECDF), found from the flashover voltages recorded during experimental testing. The distribution of best fit was then analysed by using the Kolmogorov-Smirnov (K-S) test, in order to determine the CDF that best represented the ECDF.

This statistical process was used in [1], applied to flow rates in hydrological studies, where gamma (Pearson type 3), lognormal, Weibull, generalised extreme value (GEV), Gumbel and normal distributions were compared using K-S statistics, as well as other methods. The objective was to find the distribution yielding the best fit to the experimental data. The K-S method was also utilised in [2], where a statistical analysis was undertaken to determine the relative contributions of the statistical and formative times to the total time to breakdown associated with discharges along a polymer-oil interface, under impulsive conditions. In [3], K-S statistics were applied to brain image analysis, where the K-S test was used to confirm the statistical significance of differences between normal and diseased images.

Therefore, due to the fluidity of the K-S statistical process, this method has been adopted here in order to facilitate a comparison between different statistical distributions, applied to experimental data on breakdown/flashover voltages of gas-solid interfaces, generated at a fixed pressure, and different levels of RH.

A. Distribution Models

Four statistical distributions, normal, lognormal, 2-parameter Weibull and 3-parameter Weibull were used in this study for analysis of the breakdown datasets. The cumulative distribution functions of the random variable, V (breakdown voltage in this paper), represented using each of the distribution models are shown in (1)-(5) [1], [4]:

Normal

$$CDF(V;\ \mu,\sigma) = \frac{1}{2}\left[1 + erf\left(\frac{V-\mu}{\sigma\sqrt{2}}\right)\right] \tag{1}$$

Lognormal

$$CDF(V;\ \mu,\sigma) = \frac{1}{2}\left[1 + erf\left(\frac{\ln V-\mu}{\sigma\sqrt{2}}\right)\right] \tag{2}$$

where the error function (*erf*) from the CDF of the normal and lognormal distribution is shown in (3):

$$erf(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-\frac{t^2}{2}} dt \tag{3}$$

2-Parameter Weibull

$$CDF(V;\ \alpha, \beta) = 1 - e^{\left(\frac{V}{\alpha}\right)^{\beta}} \tag{4}$$

3-Parameter Weibull

$$CDF(V;\ \alpha, \beta, \gamma) = 1 - e^{\left(\frac{V-\gamma}{\alpha}\right)^{\beta}} \tag{5}$$

The random variable V describes the breakdown voltage associated with each dataset, μ is the mean of the specific dataset, and σ is the standard deviation of the dataset, for normal and lognormal distributions. For Weibull distributions, α is the scale parameter, β describes the shape or gradient of the distribution, and γ is the location parameter. These theoretical CDFs were plotted alongside the practical ECDFs, and the K-S test was used to determine the optimum coefficient of determination (COD) between the various distribution models.

B. Kolmogorov-Smirnov Test

The COD test used in this paper is the Kolmogorov-Smirnov (K-S) test statistic, described in [5], which determines the best fit by the greatest vertical distance which results from the plotting of the empirical and theoretical CDFs. The rejection of the null hypothesis is achieved when the K-S test statistic is more than the critical value. The critical value used in this paper, at a p-value of 0.05, is 0.2941, which refers to the 95% confidence interval. The K-S statistic can be written as in (6):

$$D_n = \max_x |\ F_{exp}(x) - F_{obs}(x)\ | \tag{6}$$

Where D_n is the Kolmogorov-Smirnov statistic, F_{exp} is the cumulative distribution function associated with the null hypothesis, and F_{obs} is the empirical distribution from the data gathered from testing.

Shown in Fig. 1 is an illustrative example of how (6) relates to the practical examples associated with the CDF and ECDF, where the maximum distance from the ECDF (in blue) to the CDF (in red), represented by the arrow, allows D_n to be determined. The significance of this value is that it allows the largest distance that the ECDF diverges from the CDF to be determined. By plotting multiple distributions, and analysing multiple K-S critical values, the distribution of best fit can then be determined.

From using these specific functions in relation to the distribution models used, theoretical CDF data can be plotted for the normal, lognormal, 2-parameter Weibull and 3-parameter Weibull distributions, for comparison.

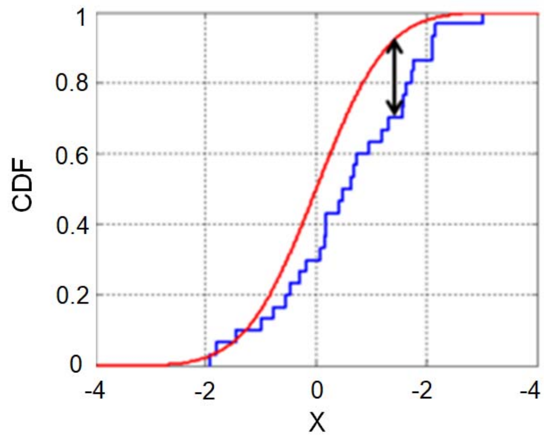

Fig. 1. Illustrative example of K-S test with the maximum point shown from the black arrow from the CDF in red and the ECDF in blue

II. Methodology

The K-S tests were conducted on experimental data generated from the testing of Delrin (Polyoxymethylene), HDPE (High Density Polyethylene) and Ultem (Polyetherimide) cylindrical spacers, situated at the center of a parallel-plane air-solid insulation system. The testing conformed to the 'step up' testing procedure in the ASTM D3426-97 standard. The flashover results were recorded for the various insulation systems, under different environmental conditions. The pressure inside the test chamber was kept constant at −0.5 bar gauge. The humidity inside the chamber was altered in order to achieve low (<10% RH), medium (~50% RH) and high (>90% RH) levels of relative humidity. All tests were conducted with a negative polarity 100/700 ns impulsive voltage, delivered from a 10-stage Marx generator. Twenty breakdown/flashover events were recorded for each test, and used in the analyses presented herein.

A. Plotting the CDF of Normal and Lognormal Distributions

The 20 breakdown voltage values generated for each test were input into an Excel spreadsheet, in order to extract the mean and standard deviation values from the dataset. The resultant mean and standard deviation generated from the distributions were plotted between 0 and 300, with iterations of 0.5, using (1) and (2), generating a normal and lognormal CDF for the specific input dataset.

B. Plotting the CDF of Weibull Distributions

The process used to enable the Weibull parameters to be extracted from the data is now described. α and β can be extracted from the data by using the intercept and slope functions on the dataset on Excel. The scale parameter, α, is equal to the 63.2% probability of failure, and the shape parameter, β, is the gradient of the resultant distribution.

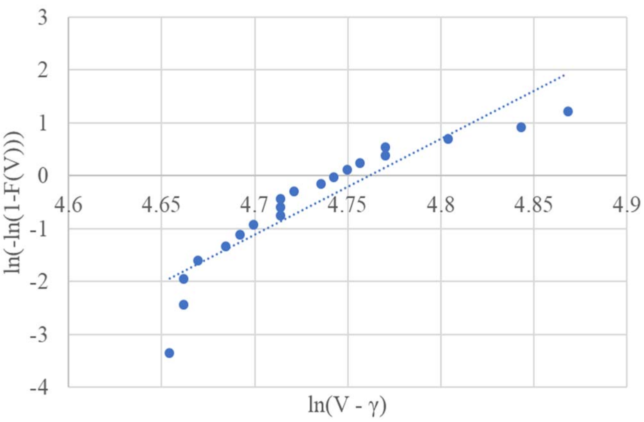

Fig. 2. Linear relationship between $\gamma = 0$ and $R = 91.21\%$ from a 2-parameter Weibull distribution

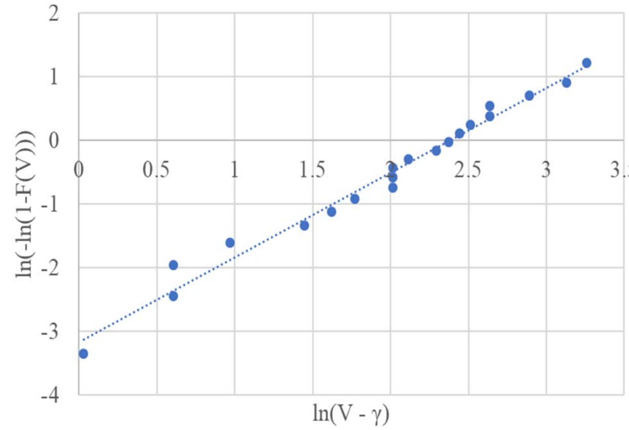

Fig. 3. Linear relationship between $\gamma = 104$ and $R = 99.09\%$ from a 3-parameter Weibull distribution

Once the data has been input, the γ value is set to 0 for a 2-parameter Weibull distribution, and therefore the α and β values can be extracted directly. An example of the resultant distribution is shown in Fig. 2. This example is for a Delrin spacer, at <10% RH.

The next step was to find the approximate γ value, for the 3-parameter Weibull distribution to be plotted. This was done by associating the γ value with the correlation coefficient, R, concluding on the best linear fit. This process was conducted by allowing multiple values for γ to be entered, therefore the relationship between the two parameters could be monitored and graphed. The R value was calculated by using the Pearson product-moment correlation coefficient for two sets of values, X and Y, represented by $ln(V-\gamma)$ and $ln(-ln(1-F(V))$ respectively, from Figs. 2 and 3, and is given by (7). Where $\bar{X}$ and $\bar{Y}$ are the sample means of the two arrays of values.

$$R = \frac{\sum(X-\bar{X})(Y-\bar{Y})}{\sqrt{\sum(X-\bar{X})^2 \sum(Y-\bar{Y})^2}} \tag{7}$$

Therefore, when the correlation coefficient (R) is maximum, the corresponding value of γ used to find this maximum value of R can be approximated as the location parameter value. This follows a method from [6], maximising the correlation coefficient in order to approximate a γ value.

Fig. 3 shows the distribution generated for the 3-parameter Weibull distribution, found by estimating γ from the maximum correlation coefficient, using (7). It can be seen that the maximum correlation coefficient of 99.09% was found when $\gamma = 104$, whereas the correlation coefficient has a value of 91.21% at $\gamma = 0$. In terms of linear regression, a better fit is found when $\gamma = 104$ (Fig. 3), compared to when $\gamma = 0$ (Fig. 2), culminating in an increase in the R value of ~8% between the distributions, when moving from $\gamma = 0$ to $\gamma = 104$. The values found using this process for each test iteration were used in order to generate the theoretical 2-parameter and 3-parameter Weibull CDFs.

C. *Combining CDF and ECDF*

In order to plot the voltages found using the ECDF in relation to each of the model CDFs, an approximation of the probability of each voltage value was calculated using the median ranks, as in (8).

$$Median\ ranks = \frac{(i-0.3)}{(n+0.4)} \tag{8}$$

where i represents the rank order (for example, $i = 1$ for the lowest breakdown voltage), and n represents the number of values in the dataset (here, $n = 20$).

The CDFs of the distribution models were then generated and are plotted alongside the ECDF in Fig. 4.

As can be taken from Fig. 4, the graph contains breakdown points where the same voltage has been found on multiple occasions, due to the resolution of the oscilloscope used during the testing process. Therefore, these points have been taken at the highest percentage found at that voltage level. These 'invalid points' have not been included on Fig. 4, to differentiate from valid points used within the K-S test. The results from the K-S test from this example used have been shown in Table I, as results meriting further discussion.

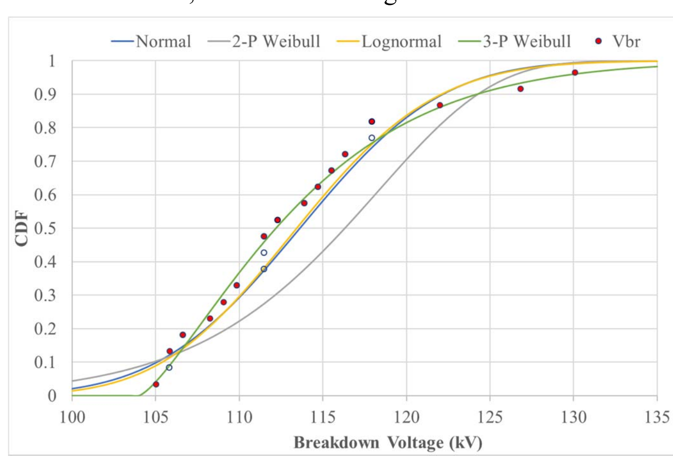

Fig. 4. Normal, lognormal, 2-parameter and 3-parameter Weibull cumulative probability functions (CDFs) with empirical distribution function found from the breakdown results; "Vbr" corresponds to the ECDF points from breakdown voltage and median rank value

IV. Results and Discussion

For all materials and environmental conditions mentioned in section II, each set of test conditions resulted in a graph similar to Fig. 4. Therefore, the K-S statistic could be deployed, in order to find the largest distance (D_n) value for each of the distributions, culminating in a best-fit conclusion. The critical K-S values found from these tests are shown in Tables I-III, where the materials are listed along with their associated critical K-S values and associated rank order of result.

TABLE I
CRITICAL K-S VALUES AND RANK ORDER ASSOCIATED WITH FLASHOVER VOLTAGE RESULTS FOR HDPE, ULTEM AND DELRIN AT –0.5 BAR GAUGE AND AT <10% RH, WITH SHADED RESULTS FROM EXAMPLE SHOWN IN FIG. 4

	Delrin		HDPE		Ultem	
	K-S critical value	Rank	K-S critical value	Rank	K-S critical value	Rank
Nor	**0.1036**	**3**	0.1228	3	0.1404	3
Log	**0.0944**	**2**	0.1074	2	0.1323	2
2- P	**0.1423**	**4**	0.1557	4	0.1599	4
3- P	**0.0665**	**1**	0.0760	1	0.0833	1

TABLE II
CRITICAL K-S VALUES AND RANK ORDER ASSOCIATED WITH FLASHOVER VOLTAGE RESULTS FOR HDPE, ULTEM AND DELRIN AT –0.5 BAR GAUGE AND AT ~50% RH

	Delrin		HDPE		Ultem	
	K-S critical value	Rank	K-S critical value	Rank	K-S critical value	Rank
Nor	0.1368	3	0.1594	3	0.1910	2
Log	0.1301	2	0.1613	4	0.1754	1
2- P	0.1401	4	0.1333	2	0.2542	4
3- P	0.1055	1	0.1207	1	0.2095	3

TABLE III
CRITICAL K-S VALUES AND RANK ORDER ASSOCIATED WITH FLASHOVER VOLTAGE RESULTS FOR HDPE, ULTEM AND DELRIN AT –0.5 BAR GAUGE AND AT >90% RH - * 3-PARAMETER WEIBULL NOT APPLICABLE AS R = MAX AT $\gamma = 0$

	Delrin		HDPE		Ultem	
	K-S critical value	Rank	K-S critical value	Rank	K-S critical value	Rank
Nor	0.1294	3	0.0566	1	0.2146	2
Log	0.1355	4	0.0574	2	0.2199	3
2- P	0.1178	2	0.1084	4	0.1612	1
3- P	0.1121	1	0.0773	3	0.1612	-*

In each table generated, the maximum K-S point (D_n) is listed for each of the 4 distributions plotted. Overall, from the rankings, the best fit to the tested data is generally the 3-parameter Weibull fit. For dry air in Table I (<10% RH), the ranking order remains consistent for each material. However, as the humidity increases, the distribution of best fit changes, where the ranking order becomes more erratic, when comparing materials at high levels of relative humidity. Yet, from the 9 tests conducted, 6 of these resulted in a 3-parameter Weibull best fit. The K-S critical value is equal to 0.2941 for $\alpha = 0.05$ [7], and the values in Tables I-III show that the D_n value was always lower than this critical value, for all tests. Therefore, for all tests, the null hypothesis is not rejected, meaning that each CDF could theoretically be used to represent the ECDF. However, the smallest critical value on average is offered by the 3-parameter Weibull distribution, which is therefore concluded as the distribution of best fit overall.

V. Conclusions And Further Work

From the analysis conducted, information has been generated on the distribution of best fit between normal, lognormal, 2-parameter Weibull and 3-parameter Weibull distributions, applied to the authors' empirical data on the breakdown/flashover voltages associated with impulsive breakdown of solid-air interfaces. From the results, a 3-parameter Weibull distribution was found to provide the best fit to the experimental breakdown voltage data, overall. As the relative humidity increased, a change in consistency of ranking orders when changing materials was witnessed.

All distributions tested, however, were able to represent each of the datasets tested in this paper, due to distance values, D_n, being smaller than the chosen critical Kolmogorov-Smirnov value.

Further work will include the statistical analysis of breakdown data generated under different conditions, such as material surface modification, polarity change, and pressure alteration. Different methods of best fit analysis, such as Anderson-Darling and Creamer-von Mises tests, will be conducted, to add to the validity of the presented results.

References

[1] Langat, Philip; Kumar, Lalit; Koech, Richard, 2019/04/09 734 - Identification of the Most Suitable Probability Distribution Models for Maximum, Minimum, and Mean Streamflow 10, 10.3390/w11040734

[2] M. J. Given *et al*., "Modifications to the von Laue statistical distribution of the times to breakdown at a polymer-oil interface," in *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 24, no. 4, pp. 2115-2122, 2017.

[3] *K. Tamaki* et al*., "Comparative study on feature descriptors for brain image analysis,"* 2014 Joint 7th International Conference on Soft Computing and Intelligent Systems (SCIS) and 15th International Symposium on Advanced Intelligent Systems (ISIS)*, Kitakyushu, 2014, pp. 679-682.*

[4] Ahmad, Mohd; Bashir, Nouruddeen; Ahmad, Hussein; M. Piah, M. Afendi; Abdul-Malek, Zulkurnain; Yusof, Fadhilah, 2013/07/01, 840, 849 Statistical Analysis of Electrical Tree Inception Voltage, Breakdown Voltage and Tree Breakdown Time Data of Unsaturated Polyester Resin, 810.5370/JEET.2013.8.4.840Journal of Electrical Engineering and Technology

[5] Chakravarti, I.M.; Laha, R.G.; Roy, J. Handbook of Methods of Applied Statistics; Wiley Series in Probability and Mathematical Statistics (USA) eng;Wiley: Hoboken, NJ, USA, 1967.

[6] Park, Chanseok - A Note on the Existence of the Location Parameter Estimate of the Three-Parameter Weibull Model Using the Weibull Plot 2018/10/09 DOI - 10.1155/2018/6056975 Mathematical Problems in Engineering

[7] L.H. Miller, "Tables of Percentage Points of Kolmogorov Statistics,"JASA,51, 1956, 111–121