21, rue d'Artois, F-75008 PARIS
http : //www.cigre.org

D1-205

CIGRE 2008

Validation of the accuracy of practical diagnostic tests for power equipment

M. BEGOVIC [1], R.N. HAMPTON [2*], R. HARTLEIN [2], J.C. HERNANDEZ-MEJIA[3], and J. PERKEL[1]
[1] Georgia Institute of Technology (ECE), Atlanta, Georgia.
[2] National Electric Energy Testing Research and Application Center, Atlanta, Georgia.
United States of America
[3] Universidad de los Andes
Venezuela

SUMMARY

Diagnostic Techniques are increasingly employed by utilities as part of integrated programmes to manage their infrastructure assets. These are sophisticated techniques being applied to complicated and diverse networks of equipment. Consequently there are many concerns that these techniques a) are not sufficiently accurate and b) could aggravate the system by, at the very least, depriving other areas of vitally short resources. Thus there is a compelling need to develop and deploy simple and robust analytical techniques that can address these problems. These evaluation approaches would then identify the effective programmes worthy of full support, while minimising the resources deployed on ineffective approaches that are ineffective.

It is not the intention of this paper to dwell on the well known issues associated with either the diagnostic techniques themselves or their detailed interpretation. Instead this paper focuses on a number of the methods we have developed to assess how well diagnostic information on cable systems relates to the performance of a specific system. Primarily this means comparing the predictions from the diagnostic information with actual service data both before and after the diagnosis. The paper looks at three main approaches:

- Direct Comparison - do assets identified as "bad" fail in service and do the "good" not fail?
- Performance Ranking - consideration of the whole continuum of performance (not just "good" and "bad") as measured by diagnostic data and correlation/validation with service experience.
- Diagnostic Outcome Mapping – how the failures in service are affected by selection, testing & maintenance actions.

The issues and experience are discussed using examples such as underground cables, poles, and transformers. The techniques described are particularly suited to the diverse nature of practical diagnostics and network architectures. In addition, the economics of diagnostic programmes are qualitatively discussed from the perspective of diagnostic accuracy and system quality disbursement.

KEYWORDS

Diagnostic-Accuracy, Reliability-Improvement, Asset-Management, Economics, Validation

nigel.hampton@neetrac.gatech.edu

I. INTRODUCTION

Utilities the world over, and especially in North America, are facing a significant future challenge to maintain and renew their assets. These ageing assets (for example, more than 20% of the presently installed underground cables are older than their design lives) are leading to ever increasing annual failures (Figure 1) whilst, at the same time, the power delivery requirements are increasing. Immediate replacement of these aged assets is not practical – the cost would be enormous and the resources required (manpower and materials) are simply not available. Thus asset management strategies are increasingly being used to help address the issue, such that the replacement of the ageing infrastructure is managed.

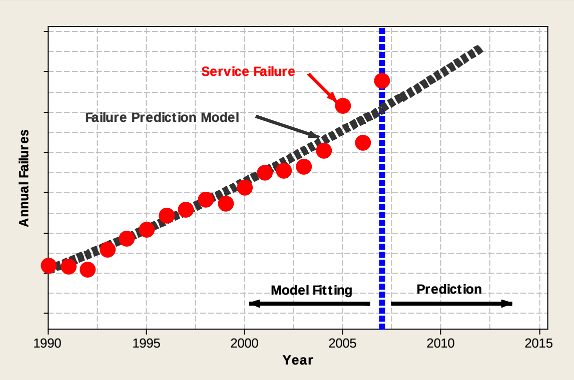

Figure 1: Increasing utility failure rates and failure prediction.

A central component of the approach to asset management is the availability of appropriate information on the assets themselves. Such information can be used to predict where the system is heading in terms of reliability (Figure 1) and to identify which assets are less reliable than others. For example, it is known that old and unjacketed underground cables are the most prone to failure and yet not every old or unjacketed cable is facing *"imminent failure"*. Thus extra information is needed if a utility is to undertake *"smart maintenance"*, that is, replacement of only those assets that will likely impact the reliability in the near future. This information is invaluable in helping to determine where maintenance and replacement funds should best be spent. Performance modeling supported by good quality and reliable diagnostic information can be a powerful tool for establishing a) the correct level of resources and b) the most effective way to use them

It is therefore clear that if we rely on diagnostic information to have an effective asset management programme, then we need to be certain that the information gathered is both relevant and accurate. We find it convenient to term this the Diagnostic Yield. In this area, most practical engineers recognise that results from diagnostic tests are not perfect (i.e. possessing accuracy near 100%). However, certain assurance is needed that the funds used to conduct diagnostic tests are well spent. They must deliver higher value compared to replacement and repair strategies based on chance selection. To this end, we have examined a number of ways to test and validate diagnostic information against the true system performance. As there are a large variety of diagnostic techniques at a utility's disposal, at least for major equipment categories, we have further concentrated on methods that are 'technique independent" and applicable to many asset types.

It is not the intention of this paper to dwell on the well known issues associated with either the diagnostic techniques themselves or their interpretation. Instead, this paper focuses on a number of the methods we have developed to assess how well diagnostic information on a particular asset relates to the performance of that asset within the system. Primarily this means comparing the predictions from the diagnostic information with system performance before and after the diagnosis. The paper will look at three main approaches:

- Direct Comparison - do the assets identified as "Bad" fail in service or, perhaps more importantly and often overlooked, do the "Good" assets not fail?
- Performance Ranking - consideration of the whole continuum of performance (not just "Good" and "Bad") as measured by diagnostic data and correlation/validation with service experience.
- Diagnostic Outcome Maps – how the failures in service are affected by selection, testing & maintenance actions.

The implications of the Diagnostic Yield upon the economic value models for Diagnostic Testing will also be discussed.

II. ECONOMIC & RELIABILITY IMPLICATIONS

The detailed economics of a diagnostic testing and action programme involve numerous interdependencies that are beyond the scope of this paper. However, when analysing the cost of maintenance, based on the results of diagnostics, it becomes very clear that two rarely considered issues are extremely important. These issues are diagnostic accuracy and system quality disbursement (i.e. what fraction of the total assets are in need of maintenance). These two components together define how much maintenance the diagnostic programme is expected to require as well as the expected number of "false negatives" the diagnostic will produce. These "false negatives" are assets that are incorrectly diagnosed as not needing maintenance when in fact they do. This type of incorrect diagnosis carries with it a consequence in the form of a service failure. Figure 2 shows how the system quality disbursement (defined by the failure rate) and diagnostic accuracy affect the percentage of "Bad" assets that are identified by the diagnostic programme. As the failure rate increases, less accuracy is needed from the diagnostic in order to identify the same percentage of assets that will produce service failures. Therefore, in a system that is in poor condition, the diagnostic does not need to be very accurate (or may even not be necessary) for a utility to be able to avoid a large number of future service failures. On the other hand, for a system that is in very good condition, the diagnostic must be very accurate for the utility to derive benefit in terms of reliability. The "false positives" will also produce a consequence but in the form of overspending on maintenance. "False positives" do not produce service failures but they do waste maintenance funds. The expected sizes of each of these groups may be computed from the diagnostic accuracy and system quality disbursement.

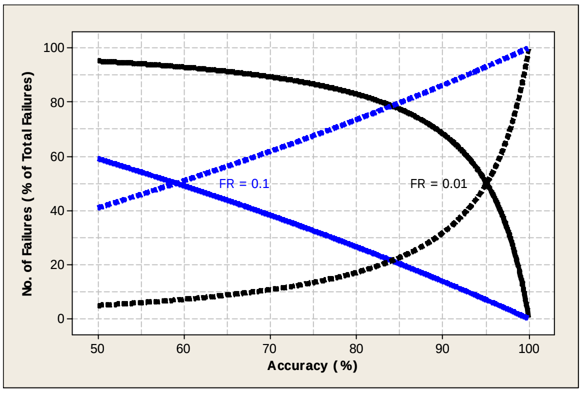

Figure 2 : Percentage of failures that are diagnosed (dashed lines) and undiagnosed (solid lines) versus diagnostic accuracy for failure rates of 0.01 and 0.1 per component per year, for a total population of 100 components.

The analyses show that the costs increase with the ratio of defective / non defective assets. The savings will accrue from the difference between the costs and the reference case. Any reference case will depend upon the specific details of the location and the Failure Tolerance of customers and regulator. Therefore, the value that can be obtained from diagnostic testing is very much related to the accuracy of the diagnostic as well as the system to which it is applied.

III. STAGES OF DIAGNOSTIC PROGRAMME

The process of employing diagnostics to increase the efficiency of reliability improvement contains four phases that are summarised as:

- Selection – Choose the assets for testing that will produce a high Diagnostic Yield. Typically this selection is based on age, failure rate, or other engineering judgment.
- Action – What actions will be performed as the result of certain diagnostic outcomes or interpretation? The actions are in two groups (Act or Not Act) and may include replacement, defer action, rejuvenation, and/or multiple levels of repair. These actions are chosen based on those that are most suitable for the system topology and most prevalent failure mechanisms (local or global defects).
- Generation – Diagnostic tests generate data that are well fitted to the type of maintenance actions and prevalent failure mechanisms.
- Evaluation – Are the methods employed for Selection, Action, and Generation, giving the expected results: lower rates of failure / increased times between failure? Can the diagnostic be improved?

Evaluation may appear to be a non-essential and abstract issue yet it has a profound practical influence. Diagnostic tests are often very costly (in the range of a few thousand dollars per day), and these monies are most often taken from the overall maintenance / replacement budget. Thus an inappropriate or ineffective diagnostic will reduce the available resources.

The focus of this paper is on the methods of Evaluation, both in terms of the accuracy of the individual diagnostic technologies (Generation) as well as the overall programme (Selection, Action, and Generation). It is important to emphasise that employing diagnostics effectively is a process that requires careful analysis and consideration before the first test is performed. The results of Evaluation depend heavily on how well this process has been conducted. This understanding / analysis begins with the data that are to be generated.

IV. DATA

Numerous asset types can be considered including transformers, rotating machines, breakers, switches, concrete or wood poles, and others. However, we will focus on our attention on one particular asset: underground cables. The analysis techniques that will be discussed in subsequent sections were developed considering the availability of data within US utilities in the period 2000 - 2006. Two types of data are needed: (1) diagnostic performance data and (2) service performance data. The level of detail contained within each of these data types is important as it limits the detail that may be obtained from any analysis to the coarsest level of the input data. For example, if diagnostic data is available for each segment of an underground feeder circuit and performance data is only available for the feeder as a whole, then the analysis is limited to looking at the feeder as a whole. This requires that the diagnostic data be analysed in such a way that the condition of the entire feeder is extracted from the condition of its individual segments. The process may be performed using a weighted average based on the relative lengths of the segments or numbers of components. As a rule, this applies to any data (diagnostic & service) that is considered for analysis. The following sections discuss the two data types needed for the validation techniques.

4.1 Diagnostic Performance Data

Diagnostic data are available in a plethora of formats and may be used in any form; but only provided it includes enough information to be able to distinguish between the assets. This becomes an issue with interpreted results that provide a class membership (i.e. this is a Level 3); as many assets may belong to the same class. We term these as "tied data". The more detailed the measurement data, the less likely ties will arise. The preference, therefore, is to have numerical measurements available even though more qualitative information can be used. The difference between the quantitative and qualitative data is in the level of interpretation needed by the analyser. It is also important to note that the diagnostic data must include a minimum of asset data in order to combine the diagnostic data with the correct failure data. The following list, though not exhaustive, summarises the suggested practical minimum information needed for analysis of underground cable assets:

- Asset Identification (typically by feeder, transformer, breaker, etc.)
- Date of test
- Asset Type/Model
- Type of diagnostic
- Numerical data
- Test protocol

4.2 Service Performance Data (Utility)

Service performance data needs to obey similar rules to that of the diagnostic data with an emphasis on the asset information. In this case, the numerical data corresponds to the number of failures before and/or after the testing. The minimum information required is as follows:

- Asset identification
- Date of failure/Age (or replacement prior to failure)
- Length or number of assets

The last category (Length / Number of Assets) becomes very important for service data as circuits typically do not experience more than a few failures; thus they are highly "tied"; for individual components like breakers or poles, this is equivalent to number per span or loop. Therefore, with some validation techniques it is more useful to look at failure rates, say failures per km or failures per component, rather than the total number of failures.

V. VALIDATION TECHNIQUES

The following sections describe the validation techniques we have developed or considered.

5.1 Direct Comparison

Direct comparison is the method that has been generally employed by workers to evaluate the effectiveness of diagnostic testing. It compares the results from diagnostic testing with the outcomes in the field by looking to see if the areas identified as "Bad" by the diagnostic actually failed within a specifc and reasonable time following the testing. In this analysis the diagnostic is assumed to only minimally affect the subsequent performance in service. This method of assessment is very onerous and typically produces overly conservative results. This effect occurs for a number of reasons:

- Method ignores assets that diagnostic shows as "Good".
- Requires that diagnostic has the ability to clearly separate assets into "Good" and "Bad" groups with no overlap. Experience shows that this condition is never fulfilled with the features / tools employed by most diagnostic tests.
- Anything less than 100% accuracy gives "Bad" results.

Notwithstanding the above, one thing is certain with a Direct Comparison: if this method shows that things are, in fact, "Good" then the reality is that the diagnostic has done an excellent job of identifying the "Bad" components. On the other hand, if Direct Comparison indicates the diagnostic was not effective then other methods should be employed to evaluate the diagnostic's performance as this method tends to exaggerate any imperfections.

5.2 Performance Ranking

This technique uses the same basic data as the Direct Comparison approach, and has been developed in our group as a means of evaluating the effectiveness of diagnostic testing by comparing the diagnostic data with actual system performance. This comparison provides measures, quantitative & semi quantitative, of the accuracy of the diagnostic. Performance Ranking is the only technique that maps the data all the way from the best to the worst. It is not focused on the "Bad" assets as in the case of Direct Comparison. In addition, it may be used with any diagnostic test was well as with data provided in any form more detailed than a simple pass/fail. Example diagnostics include (but are not limited to) Partial Discharge (offline and online), Tan δ, DGA, pole inspections, etc. An example of Performance Ranking is shown in Table I.

Table I: Illustrative Performance and Diagnostic (*A, B & C*) for 5 assets.

Asset **Age**	**I** **18 Yrs**	**II** **29 Yrs**	**III** **16 Yrs**	**IV** **18 Yrs**	**V** **15 Yrs**
Failures in Service	1	2	1	0	1
Performance Rank	**4**	**1**	**3**	**5**	**2**
Diagnostic *A – level based*	0	5	4	2	2
Diagnostic *A* Rank	**5**	**1**	**2**	**4**	**3**
Diagnostic *B – value based*	5	20	5	10	22
Diagnostic *B* Rank	**5**	**2**	**4**	**3**	**1**
Diagnostic *C – Amount of Refurbishment*	2%	15%	0%	15%	20%
Diagnostic *C* Rank	**4**	**3**	**5**	**2**	**1**

Performance Ranking is completed by generating two distinct ranks (a number representing 1st, 2nd, 3rd, etc), Performance Rank and Diagnostic Rank, for each tested asset. Each of these ranks is a number that gives the relative performance of each asset as it compares to all other assets in the group. There cannot be duplicate ranks within either rank type. Furthermore, all circuits must be assigned both a Performance and Diagnostic rank for plotting. In other words, if a test group consists of 10 segments then there will be at most one #1, one #2, one #3, etc., for the Performance Rank and then the same would hold true for the Diagnostic Rank, i.e. they should both map onto a set {1,2,3,4,5,6,7,8,9,10}. The basic procedure can be summarised as follows:

1. Determine the "Diagnostic Rank" using the available diagnostic data and the asset information.
2. Determine the "Performance Rank" using the available failure and asset information.

3. Plot Diagnostic Rank versus Performance Rank.
4. Determine the best fit line with an appropriate statistical method.

The concept of ranking the circuits is quite simple. However, with test groups containing more than a few assets it is very likely that there will be cases where the ranking criteria produce ties (Table I). As one of the requirements of this technique is to assign a single rank to each circuit, breaking these ties becomes critical. Several methods have been developed to address this issue for both ranks. Each will be discussed in detail in the following sections in conjunction with the steps outlined above.

5.2.1 Performance Rank

The Performance Rank is based on the failure data from either before or after testing. It is determined by comparing the failure rates (annual or cumulative) for all tested components with one another and then ranking from worst (highest failure rate) to best (lowest failure rate). In the case of cables, the ranking is most commonly done considering multiple segments together as one "feeder" as this is the extent of detail available in the failure records at most utilities. This is not an issue for other devices such as transformers, breakers, or poles. However, it is important to note that the ranking approach is able to cope with whatever detail is available in the data.

5.2.2 Diagnostic Rank

The Diagnostic Rank is far more complicated to determine than the Performance Rank as different diagnostic techniques provide their assessments in different ways. The data may be quantitative measures of the degradation that has occurred in the device or may simply be qualitative such as "Good", "Bad", or "Okay" etc. Furthermore, this data may be as specific as by component or may include several components at once. Whatever the level of detail may be, it is necessary to evaluate the diagnostic data in the same groupings as the performance data. An example would be, two cable segments may have been tested separately, however, they must be considered as one because the failure records do not distinguish between them.

It must be emphasised that the only requirement for diagnostic data is that it be capable of providing some level of distinction between different circuits.

5.2.3 Ranking Tie Breaks

As mentioned above, it is common to see situations (Table I) where ties can arise, especially in the case of the Performance Rank. These cases can be solved in a number of ways depending on the type of component. For example, components such as breakers and transformers may be handled by considering age or even the number of exposures to fault currents whereas cables can be dealt with by considering the length. Whatever the asset, it is important to choose a characteristic that will include sufficient variability within the asset population. It is also possible, however, that multiple characteristics will be needed in order to break all the ties. Therefore, it is possible to rank any dataset, however, the level of interpretation differs depending on the number of characteristics that are needed in order to break the ties.

The hierarchy was developed for cables using the circuit information that is typically available at utilities:

- Measurement data (Diagnostic Rank).
- Circuit length: Average per unit length. Also, longer circuits should be more prone to failure so give higher rank to longer circuits – Table I.
- Number of accessories: More accessories lead to more opportunities for failure so give higher rank to circuits with more accessories.
- Age: Older circuits receive a higher rank as these are logically more prone failure.
- Construction: Primarily, insulation type, however, this should also include jacketing, whether the cable was direct-buried or installed in conduit, and type of neutral.

5.2.4 Analysing the Ranks

Once the two ranks have been computed they may be analysed either graphically (qualitatively) or statistically (quantitatively). In the former case, a plot of Diagnostic Rank versus Performance Rank is generated. A sample of such a plot is shown in Figure 3 which uses real diagnostic tests using different approaches on a cable system.

The accuracy of the diagnostic test is directly related to how far from the dashed line the dots are in Figure 3. The interpretation of Figure 3 is as follows: Circuits in the lower left corner, we consider by convention to be the worst performers (highest failure rate and classified as "Bad" by the diagnostic test) while the upper right corner contains the best performers (low failure rate and classified as "Good" by the diagnostic test).

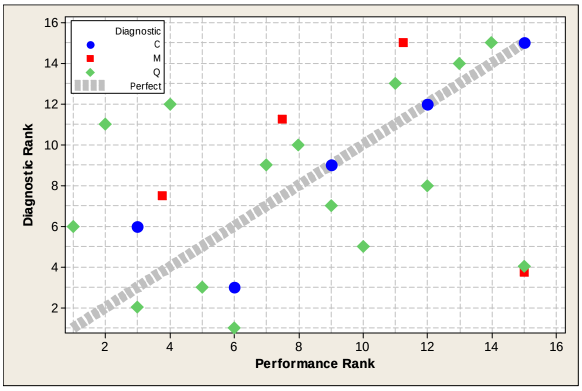

Figure 3: Sample Performance Ranking plot. Dots represent individual segments whereas the dashed line would pass throught the points representing the perfect correlation between Performance and Diagnostic ranks.

The dashed line can also be thought of as the perfect correlation between the performance and diagnostic ranks. Therefore, the obvious statistical approach is to examine the Pearson Correlation Coefficient [3], [4] between the two ranks as shown in (1).

$$r_{DP} = \frac{n\sum D_i P_i - \sum D_i \sum P_i}{\sqrt{n\sum D_i^2 - \left(\sum D_i\right)^2}\sqrt{n\sum P_i^2 - \left(\sum P_i\right)^2}} \quad \textbf{(1)}$$

Where, r_{DP} = Pearson correlation coefficient, n = number of samples, D_i = i^{th} Diagnostic Rank, P_i = i^{th} Performance Rank

The value of the correlation coefficient is in the range [-1,1] where one corresponds to a perfect correlation, zero to no correlation, and negative one to an inverse correlation. Comparison of these coefficients enables the computation of the difference in accuracies between different diagnostic technologies. Table II shows the Pearson correlation coefficients for the data shown in Figure 3. The respective significance levels indicate that only the results from Diagnostic C are unlikely to occur randomly (with probability <0.05).

Table II: Pearson correlation coefficients for three diagnostic techniques.

DiagnosticTechnique	r_{DP}	**Level of Significance**
C	**0.900**	**<0.05**
M	**-0.200**	**>0.1**
Q	**0.321**	**>0.1**

Figure 3 deals with service data using 3 diagnostics, the example described in Figure 4, looks at field-aged underground XLPE cable samples in which Tan δ was measured at 0.1 Hz. using a sinusoidal Very Low Frequency (VLF) source at U_0 and then taken to failure (breakdown) using a VLF (0.1 Hz) step ramp protocol. Figure 4 shows the breakdown voltage versus Tan δ value for each of the samples. Note that the samples shown in blue failed during 60 Hz. Tan δ measurement.

Direct inspection of Figure 4 is not a straightforward way of understanding the manner in which the Tan δ value predicts the breakdown strengths of the samples. However, if the data are analysed using Performance Ranking, the picture becomes much clearer. Figure 5 shows the resulting Performance Ranking plot for this data. Computation of the Pearson Correlation coefficient for the six tested samples yields a value of 0.967 which is significant at the 0.001 level. Therefore, these results would only occur randomly with a probability of less than 0.1%. Any strength test may be analysed in this manner including those related to poles, breakers, switches, surge arresters, cutouts, etc.

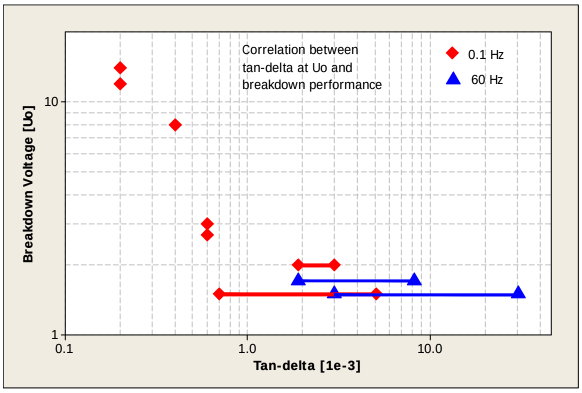

Figure 4: VLF breakdown strength versus average Tan δ value for service aged cables measured in the laboratory.

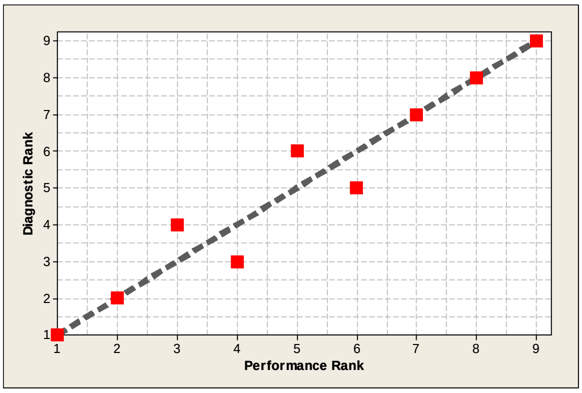

Figure 5: Performance Ranking plot for Tan δ and breakdown strength for data from Figure 4.

5.3 Diagnostic Outcome Mapping (DOM)

DOM uses failure and testing data to evaluate whether or not changes in failure rate are coincident with diagnostic activities (as well as the activity called for by the test results). DOM uses a graphical representation for analysis. Experience has shown that the well established Reliability Growth Model (often referred to as Crow-AMSAA technique specified in IEC 61164 [2]) is very suitable. Crow-AMSAA is a plotting technique that plots cumulative failures versus time on log-log scales. The instantaneous failure rate is determined by computing the slope, or gradient, of the curve at any particular point. A decreasing gradient indicates the failure rate is decreasing while an increasing gradient corresponds to an increasing failure rate. By adding the testing events to the same representation, it is possible to determine the effect that the magnitude of the test programme (diagnostic testing plus required action) has had on the reliability.

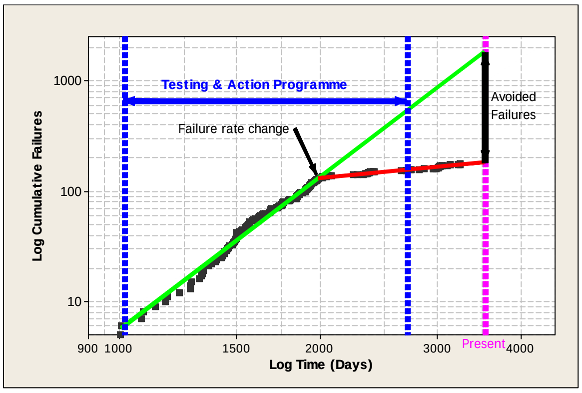

Figure 6: A utility outcome map in which the failure rate decreases as a result of the diagnostic testing and action programme.

Figure 6 shows one of the possible outcomes that can occur for a particular underground cable system. In Figure 6 it is seen that the testing and action programme eventually produced a significant reduction in failure rate at approximately 2000 days. The alteration in gradients on the DOM plot shows this change in the failure rates. It is important to note that this change is not immediate. It requires a significant commitment on the part of the utility before positive results like those shown in Figure 6 will be observable. At present, it is estimated that the utility was able to avoid approximately 900 failures through its programme. The reduction in failure rates continues after the program has been halted and is evidence of the persistence of beneftis that a diagnostic programme can provide.

It is also possible and quite common to see constant rates and increasing failure rates following testing events for individual circuits and then even for an entire system. Figure 7 shows another example of an overall system outcome map since the initiation of a diagnostic testing and action programme. This approach extends the analysis in Figure 6 from a single circuit to a larger at risk population.

In Figure 7 it is clear that there has been little change during the first three years of the programme; however, the programme yields a reduced failure rate, seen by the depression of the data from the linear fit line, during the last three years. This observation may be investigated further by examination of the annual gradients for the curve in Figure 7. Comparison with the amount of testing and action accomplished within a particular year provides the results shown in Figure 8. This figure clearly shows the annual gradients increase out to year 3. This is presumed to represent the continuation of the pre programme trend. The failure rates then began to decrease after year three. That year corresponds to the point at which the number of tests and actions exceeded a critical level, approximately 175 actions, where the programme was able to bring the failure rate under control. This analysis illustrates the usefulness of DOM in showing real benefits of a diagnostic testing and action programme. It also shows that for a programme to make a real impact on the system reliability it must reach a minimum level of activity. This level will differ from system to system and will require a long term commitment on the part of the utility if successful results are to be achieved.

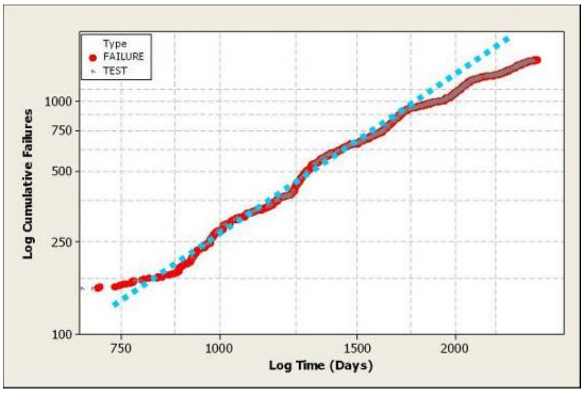

Figure 7: Sample system outcome map that shows reduced failure rate.

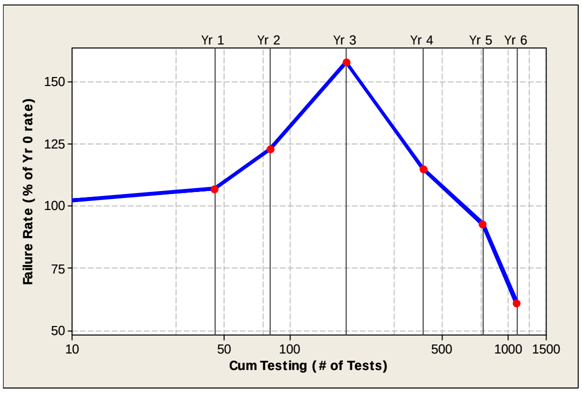

Figure 8: Failure rate versus cumulative tests for an overall testing and action programme.

VI. CONCLUSIONS / FUTURE WORK

This paper describes two primary techniques for evaluating the accuracy and effectiveness of diagnostic techniques and utility programmes for employing them. It is clear that the accuracy of a diagnostic and system quality disbursement will greatly influence the value that a utility can hope to obtain. Furthermore, utilities must be aware that this value also depends heavily on their understanding of their assets and ability to select and act in ways that take advantage of what diagnostics can offer. The enitre process takes considerable care in planning and patience as positive results accrue slowly after inception of the program and the only way to assess the benefits is through Evaluation.

VII. ACKNOWLEDGEMENTS

The authors gratefully acknowledge the useful discussions with many of the engineers involved within the Cable Diagnostic Focused Initiative. The work reported here was supported by a large number of utilities in North America and the US Department of Energy under award number DE-FC02-04CH11237

BIBLIOGRAPHY

[1] IEEE Power Engineering Society, "IEEE Guide for Field Testing and Evaluation of the Insulation of Shielded Power Cable Systems," [IEEE Std. 400-2001].

[2] International Electrotechnical Commission, "Reliability Growth – Statistical Test and Estimation Methods," [IEC 61164, Second Edition 2004-03].

[3] E.W. Weisstein, "Correlation Coefficient" [From *MathWorld*, Wolfram Web Resource. http://mathworld.wolfram.com/CorrelationCoefficient.html].

[4] G.E.P. Box, W.G. Hunter, J.S. Hunter, *Statistics for Experimenters*, [John Wiley & Sons, Inc., New York, NY, USA, 1978, 60-91].