Equipment Failure Forecasting Based on Past Failure Performance and Development of Replacement Strategies

Miroslav Begovic[a], Joshua Perkel, and Rick Hartlein
School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta GA 30332, USA

[a]E-mail: miroslav@ece.gatech.edu

(Received October 12 2006, Accepted October 19 2006)

When only partial information is available about equipment failures (installation date and amount, as well as failure and replacement rates), data on sufficiently large number of yearly populations of the components can be combined, and estimation of model parameters may be possible. The parametric models may then be used for forecasting of the system's short term future failure and for formulation of replacement strategies. We employ the Weibull distribution and show how we estimate its parameters from past failure data. Using Monte Carlo simulations, it is possible to assess confidence ranges of the forecasted component performance data.

Keywords : Falure prediction, Weibull distribution, Replacement scheduling

1. INTRODUCTION

The problem of resource management has long been recognized as one of the burning issues in electric utilities. Knowing how much to invest in creating a reliable and successfully performing resource pool (i.e. distribution cable network), when to repair or replace, and what human and financial resources are needed from year to year in order for such a network to operate successfully, the answers to those questions may represent substantial savings for the utility. Among the most acute problems that utilities are facing is the problem of accurate logging of system past performance and failure rate. As far as cables go, very little or no information is available to support such an activity.

The purpose of the paper is to propose a methodology for identification of failure performance of cable populations when only incomplete past failure data is available. It is assumed that a population of cables of the same type is being tracked over a period of several years, with knowledge of number of installed miles per year, number of replaced miles per year, and failures in any given year is available. The age of the cables which fail is not assumed to be known, as it is commonly the case in the databases maintained by the utilities.

Prior work by one of cable reliability researchers (Bill Forrest,[8]) is used here as a basis to extract the parameters of the Weibull distribution, which is assumed to describe the failure rate performance of the entire cable population. We have expanded and modified that approach to include multiple parameters identification and nonlinear models, and tested it using field data which was obtained from an actual cable population.

In addition, the have expanded the methodology to include Monte Carlo simulations of the failure rates in order to produce the estimates of distributions of failures rather than the most likely estimates. By doing so, we have developed a capability to associate confidence ranges with estimates of failure and replacement rates that are forecasted in the short time horizon of one to three years into the future. By doing so, planning can be associated with the desired level of confidence, which provides better quality information for a cost-conscious utility planner. It should be noted that the accuracy of the proposed methodology strongly depends on the quality and quantity of the input data, and it is envisioned that it could be enhanced in the future by combining chronological failure rate information with some form of condition monitoring, which can be coupled with the failure model that may sharpen the accuracy of the failure forecasts needed for a precise planning.

If a large amount of statistical information about component failure rate performance is available (which it is usually not), then accurate statistical models can be

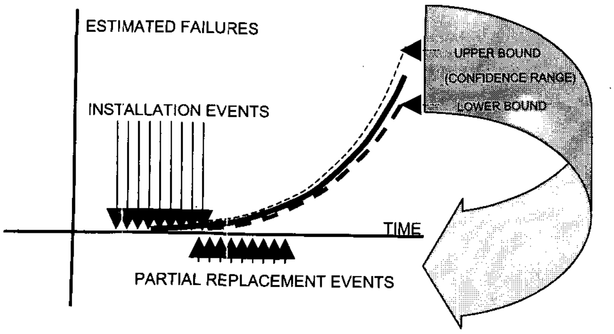

Fig. 1. Illustration of replacement scheduling events based on desired failure performance.

designed around them and used for predictive maintenance strategy development[4,5,7]. Often times, only partial information is available: we assume here that a database of past failures contains only the following information: year of installation and number of components, amount of components replaced in any given year, and the total number of failures in any year. It is not assumed to be known the age of failed components, as such statistics are rarely known in the utilities. By combining a large number of yearly installations of the components (which must be of the same type for consistency of statistics), estimation of Weibull parameters[8] may be accomplished, from which forecasting into the short term future performance may be possible.

2. PROBLEM STATEMENT

Suppose that $p(t)$ is the probability density function (PDF) of the time to failure t of a single component. The probability of that component failing before time t is given by

$$P(t) = \int_0^t p(u)du. \tag{1}$$

If we have a system of N such components connected in series, the probability that the system will fail is

$$P_0(t) = 1-(1-P(t))^N \tag{2}$$

where $P_0(t)$ is cumulative distribution function (CDF) of the time to failure, and the corresponding PDF is

$$p_0(t) = N(1-P(t))^{N-1}p(t). \tag{3}$$

For example, if $p(t) = e^{-t}$, $t \geq 0$, and $p(t) = 0$, $t < 0$; and let $N = 2$, then

$$p_0(t) = 2(1-1+e^{-t})e^{-t} = 2e^{-2t}, \quad t \geq 0. \tag{4}$$

For this distribution, the expected time to failure of a single component is $E(T) = 1$ and for the system with two components, $E(T_o) = 1/2$ (this is not true in general (see below)).

The Weibull distribution [1,2,5] is arguably the most popular parametric family for modeling reliability and survival phenomena. If the times to failure T are distributed as Weibull $Wei(\alpha,\beta)$, with parameters α - scale parameter and β - shape parameter, the PDF has the form

$$p(t) = \beta\alpha^{-\beta}t^{\beta-1}e^{-\left(\frac{t}{\alpha}\right)^\beta}, \quad t \geq 0 \tag{5}$$

its CDF is

$$P(t) = 1-e^{-\left(\frac{t}{\alpha}\right)^\beta}, \quad t > 0 \tag{6}$$

and the PDF of time to failure of the overall system is

$$p_0(t) = N\beta\alpha^{-\beta}t^{\beta-1}e^{-(t/\alpha)^\beta N}, \quad t \geq 0. \tag{7}$$

Depending on the values of its parameters, the Weibull distribution can model a range of different reliability behaviors. For example, the value of the shape parameter β dictates the behavior of failure rate function. If $\beta>1$ ($\beta<1$), the failure rate is increasing (decreasing) with time, while for $\beta=1$ the Weibull distribution coincides with the exponential distribution and, as it is well known, the failure rate is constant in time in such a case.

It is interesting to observe that $p_0(t)$ can be rewritten as

$$p_0(t) = \beta\alpha_m^{-\beta}t^{\beta-1}e^{-(t/\alpha_m)^\beta}, \quad t \geq 0 \tag{8}$$

where

$$\alpha_m = \frac{\alpha}{N^{1/\beta}} \tag{9}$$

which means that $p_0(t)$ is also a Weibull density. The expected value of T (the time to failure of a single component) is

$$E(T) = \alpha\Gamma\left(\frac{1}{\beta}+1\right) \tag{10}$$

and the expected value of the time to failure of a system consisting of N identical units is

$$E(T_0) = \frac{\alpha}{N^{1/\beta}}\Gamma\left(\frac{1}{\beta}+1\right). \tag{11}$$

When α = β = 1 (which is the case of the exponential distribution) we get

$$E(T_0) = \frac{E(T)}{N}. \tag{12}$$

Otherwise, obviously, this relationship will not hold. For example if α = 1 and β = 2, the relationship is

$$E(T_0) = \frac{E(T)}{\sqrt{N}}. \tag{13}$$

Given PDF and CDF, the failure rate is defined by

$$f(t) = \frac{p(t)}{1-P(t)}, \quad t \geq 0 \tag{14}$$

and for the Weibull distribution it is

$$f(t) = \beta\alpha^{-\beta}t^{\beta-1}, \quad t \geq 0. \tag{15}$$

For the overall failure rate we have

$$f_0(t) = \frac{p_0(t)}{1-P_0(t)}, \quad t \geq 0 \tag{16}$$

which in fact is of the same form as that of $f(t)$ except that instead of α we use α_m, i.e,

$$f_0(t) = \beta\alpha_m^{-\beta}t^{\beta-1}, \quad t \geq 0. \tag{17}$$

In terms of the original α and N, we have

$$f_0(t) = N\beta\alpha^{-\beta}t^{\beta-1}, \quad t \geq 0. \tag{18}$$

We therefore begin with the assumption that the expected number of failures that occur in a population of X components of the same type at time t years after the installation is given by

$$N_f(t) = X \cdot a \cdot (t-g)^b, \quad t > g. \tag{19}$$

where a is a scaling constant, b is a constant which is related to time dependency, and g is a quiet period (without failures) following the initial deployment of the component. If a component is installed in year i following the first installation and consists of X_i units, then the expected number of failures at t years after the *initial* population installation will be:

$$N_f(t) = X \cdot a \cdot (t-g-i)^b \text{ for } t > g+i. \tag{20}$$

Under the assumptions used in the above derivation (Weibull distribution), the failure rate possesses a linear relationship with the number of components. Finally, if we combine component populations installed in years *1, 2,...,i, i+1, ..., n*, the cumulative estimated (in some sense, most likely) number of failures of such a population will be[8]

$$F(a,b,t,g) = \sum_{i=1}^{n} N_{f_i}(t) = \sum_{i=1}^{n} X_i \cdot a \cdot (t-g-i)^b \quad \text{for } t > g+i \tag{21}$$

That is a four-parameter function of time. Our objective is to identify the three unknown parameters (a,b,g) from the knowledge of the observed number of failures over a finite (often quite short) period of time, by extracting the needed parameters from the observations by fitting the model to the observations in the least squares sense. Let us describe the problem in terms of the following quantities: if X_i is the number of installed components in year i ($i \in \{1,2,...,n\}$), the population installed in year i will experience $h(X_i,t)$ failures in year t,

$$h(X_i,t) = X_i \cdot a \cdot (t-g-i-1) \cdot u(t-g-i-1) \tag{22}$$

where u is a step function of time

$$u(t) = \begin{cases} 0, & \text{if } t < 0 \\ 1, & \text{if } t \geq 0 \end{cases}. \tag{23}$$

The function $u(t)$ facilitates implementation of the "zero failures before time g" rule. The total number of failures in year t will be equal to the sum of failures of all populations (assuming, for a moment, that the time exponent $b = 1$):

$$H(\Sigma X,t) = \sum_{i=1}^{n} h(X_i,t) = \sum_{i=1}^{n} X_i \cdot a \cdot (t-g-i-1) \cdot u(t-g-i-1). \tag{24}$$

For the sake of simplicity, we can start counting years from $g+1$ (the first year when a non-zero number of failures is expected to occur) so that (after the change of time reference) year number $g+1$ becomes year 1, year $g+2$ becomes year 2, etc. Let us now assume that the actual observed numbers of failures in years 1, 2, etc., are f_1, f_2, etc. respectively The difference between estimated and observed numbers of failures in year i is

$$\Delta_i = F_i - f_i. \tag{25}$$

We form now the sum of squares of differences Δ_i for all years 1 through n:

$$\Delta = \sum_{i=1}^{n} \Delta_i^2 = \sum_{i=1}^{n} (F_i - f_i)^2. \tag{26}$$

The only unknown in the above expression is the parameter a, which represents the population-dependent constant we would like to determine. We calculate the value of a that minimizes Δ. This minimum is reached when $\partial\Delta(a)/\partial a = 0$,

$$\frac{\partial\Delta(a)}{\partial a} = \sum_{i=1}^{n} 2\cdot(F_i(a)-f_i)\cdot\frac{\partial F_i(a)}{\partial a}$$
$$\Rightarrow \sum_{i=1}^{n} f_i \cdot \frac{\partial F_i(a)}{\partial a} - \sum_{i=1}^{n} F_i(a)\cdot\frac{\partial F_i(a)}{\partial a} = 0. \tag{27}$$

Solution to

$$\frac{\partial\Delta(a)}{\partial a} = \sum_{i=1}^{n} f_i \cdot \frac{\partial F_i(a)}{\partial a} - \sum_{i=1}^{n} F_i \cdot \frac{\partial F_i(a)}{\partial a} = 0 \tag{28}$$

will yield the optimal value of a, which will be denoted $\hat{a}$, i.e.,

$$\frac{\partial\Delta(\hat{a})}{\partial a} = \sum_{i=1}^{n} f_i \cdot \frac{\partial F_i(a)}{\partial a} - \sum_{i=1}^{n} F_i \cdot \frac{\partial F_i(a)}{\partial a} = 0. \tag{29}$$

In the equations so far, we have assumed nothing about removing portions of components from service. In order to develop the algorithm for determining the elements of matrix R_j (removed components), we shall assume that the following is known: X_i, $i = 1,2,\ldots,n$ are numbers of installed miles from year 1 until year n. Let the vector $R = [r_1\ r_2 \ldots r_n \ldots r_{n+k}]^T$ represent the quantities of cable removed from service in the years $1, 2, 3, \ldots, n, \ldots, n+k$ respectively. In order to present the algorithm, we shall assume that the following is known: X_i, $i = 1,2,\ldots,n$ – number of installed components from year 1 until year n; X_j, $j = n+1,n+2,\ldots,n+k$ - number of installed components from year $n+1$ until year $n+k$; R_i, $i = 1,2,\ldots,n$ - number of removed components from year 1 until year n.

Our objective is to find the numbers R_j, $j = n+1, n+2,\ldots, n+k$ that represent quantities of components to be removed in the period from year $n+1$ until year $n+k$. If we combine equipment populations installed in years $1, 2, \ldots, i, i+1, \ldots, n$, the number of yearly failures of such a population will be (assuming that b is different from 1, and adjusting the matrix elements accordingly):

$$\begin{vmatrix} F_1 \\ F_2 \\ \vdots \\ F_{n-1} \\ F_n \end{vmatrix} = a \cdot \begin{vmatrix} 1^b & 0 & \cdots & \cdots & 0 \\ 2^b & 1^b & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ (n-1)^b & (n-2)^b & \cdots & 1^b & 0 \\ n^b & (n-1)^b & \cdots & 2^b & 1^b \end{vmatrix} \cdot \begin{vmatrix} X_1 \\ X_2 \\ \vdots \\ X_{n-1} \\ X_n \end{vmatrix}. \tag{30}$$

As has been said, we would ideally like to know *which* yearly populations have been affected by a removal in any given year, but such knowledge is for the moment assumed not to be available. For lack of better information, we assume that any removal of components from service occurs on the oldest vintage still available and in service.

Let us define the matrix X in the following way:

$$X = [X_{ij}]_{n\times n} \tag{31}$$

where X_{ij} represents the amount of components, installed in year #i and remaining in service in year #j. Also, assume that the vector F_j, $j = n+1,n+2,\ldots,n+k$ represents the estimated number of failures in years $n+1$, $n+2$, $n+3$, $\ldots$, $n+k$, respectively.

As we know the elements of X up to the time when all installations and replacements are known, the solution of the equations yields the parameter set $\{a,b\}$. With knowledge of the parameters, a set of equations can be solved for any desired time horizon $\{n+1, \ldots, n+k\}$ in order to determine: **i)** the estimated number of failures when a replacement schedule is planned for and known in that period; or **ii)** the estimated necessary replacement schedule, which should maintain the estimated number of failures at the desired (planned) rate within the time horizon of interest. In practical terms, the time horizon should be as short as reasonably possible in order to avoid the accumulation of uncertainty that would invalidate the results.

3. ILLUSTRATION OF THE PROPOSED METHOD

A synthesized data set was used for simulations. It contains information on numbers of installed and replaced components, as well as the number of observed failures over a period of 33 years. The data set was modified from an experimental data set for power system distribution cables. The compliance of the data set with Weibull distribution is evidenced from Fig. 2, which indicates a good fit.

The results from Table 2 are compiled into Fig. 4, where the actual numbers of yearly component failures (indicated by labels '+') are superimposed to the solid line, showing the estimated numbers of component failures obtained using the proposed algorithm. The upper dashed line represents the estimation of the numbers of yearly failures which would have been experienced by the component population had the partial yearly replacements not been applied as per Table 1. The net reduction of estimated failures is the result of younger component population (all replacements are assumed to substitute the oldest components in service at the time of replacements).

It is desirable to extend our procedure using probabilistic simulation. That would create the opportunity to analyze the *probability distributions* of failures rather than their *estimated values*). We shall employ Monte Carlo technique. The basic procedure for obtaining s predictions from a simulated dataset, n years long, is as follows:

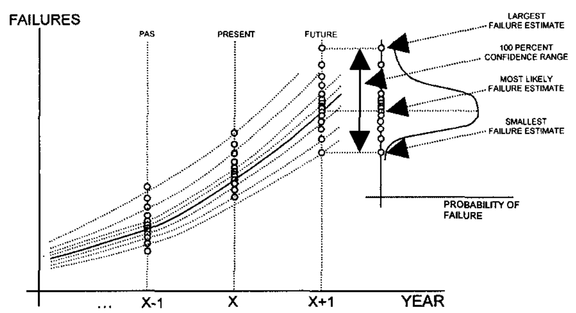

Fig. 2. Use of Monte Carlo simulations for failure forecasting into a short term time horizon.

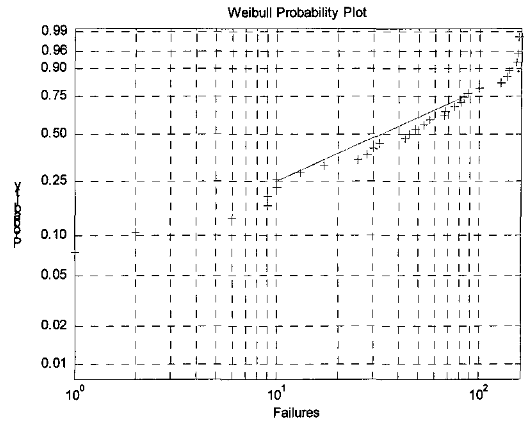

Fig. 3. Weibull probability plot of failure data.

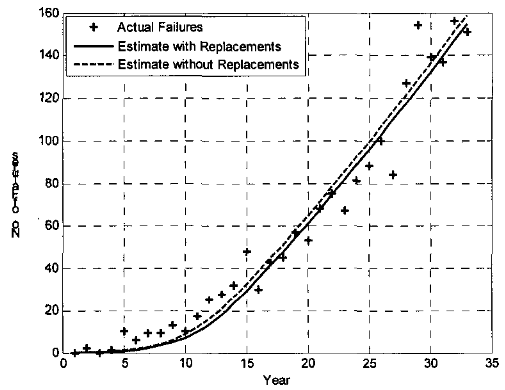

Fig. 4. Estimated number of component failures assuming no replacements (upper curve) and replacements (lower curve) superimposed over the actual numbers of component failures (shown as points labeled '+').

1. Determine the Weibull parameters from the primary dataset (chronological failure data) using the procedure already developed.
2. Use the identified parameters to determine the points on the optimized fit curve (estimated numbers of failures) for all years up to year n.
3. Generate failure distributions, according to the Weibull distribution identified for the entire data set and the estimated number of failures for each year; each random sample of synthesized failures should contain s random samples.
4. Construct s datasets by selecting randomly one sample from each of the n yearly failure distributions.
5. Estimate the number of failures and/or replacements using the procedure developed earlier for each of the new s datasets.

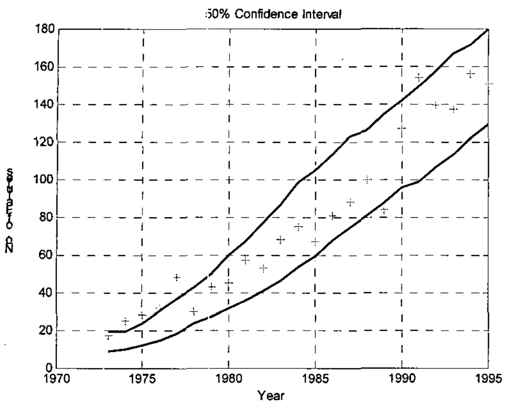

Fig. 5. 50 % confidence interval for estimated failures. Symbols '+' represent the actual observation data. The original data set is shown in Table 2.

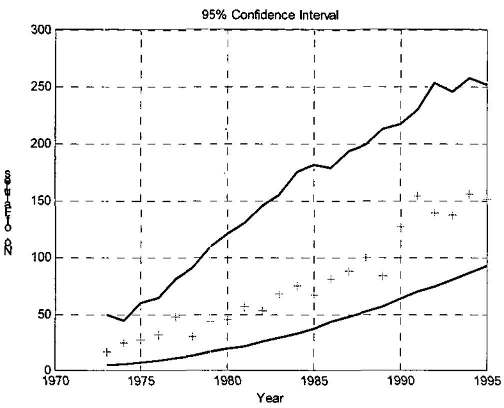

Fig. 6. 95 % confidence interval for estimated failures. Symbols '+' represent the actual observation data. The original data set is shown in Table 2.

Use the results of s simulations (estimated failures) to calculate the distribution of estimated failures, as well as the impact of assumed component replacement rates on distribution of estimated failure rates, or distribution of estimated replacement rates necessary to result in a desired failure rate in the future. Figures 5 and 6 illustrate the confidence ranges (50 % and 95 %) obtained by running a Monte Carlo simulation on the same data set (described in Table 2).

Figure 6 shows the relationship between upper bound confidence levels and corresponding replacement rates.For the system planners, such a curve (obtained for any desired failure scenario) could be the resource from which to obtain the final answers to the questions like: "How many feet of cable should be replaced so that in year 1 the estimated number of failures does not exceed X with probability Y?" That way, the notion of risk is included in the decision making process.

4. CONCLUSION

The algorithm presented in this paper relies solely on basic historical data to forecast the number of failures. As a consequence of the available data, several assumptions were made from the beginning to make the analysis possible. These assumptions are:

- The components have a lifetime consistent with a three-parameter Weibull distribution.
- The actual component that failed is unknown so it is assumed that the oldest components are always replaced first.

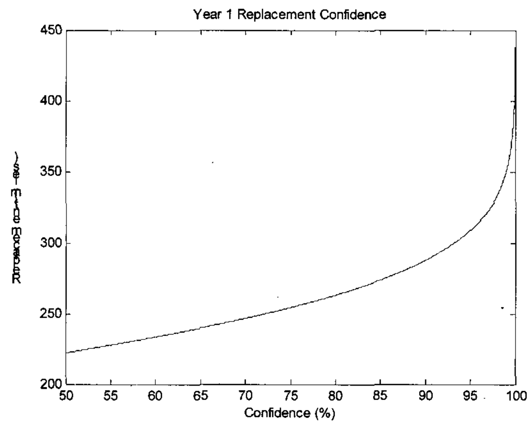

Fig. 7. The relationship between upper bound confidence levels and replacement estimates for year 1 predictive replacements necessary to maintain a constant failure rate.

In addition, the algorithm allows for repeated calculation of failures assuming the component population has been altered. This allows for the calculation of replacement components needed to achieve desired changes in the failure curve. In effect, the algorithm can predict how actions in the present will impact the overall failure trend.

On the other hand, a single prediction is virtually meaningless as there is no certainty attached to it. Monte Carlo simulation techniques have been employed to extract some level of certainty. By performing thousands of simulations, the algorithm yields a distribution for each of the parameters of interest and from these confidence intervals may be extracted. These tell how likely the forecasted parameter is to fall within a certain range.

This algorithm may be extended to include data obtained through condition monitoring to increase the accuracy of results. It may also be modified to take advantage of more complete historical data thereby eliminating one of the necessary assumptions. A final modification would be to dispose of the assumed Weibull distribution and use non-parametric methods to generate the predictions. This final suggestion would require significant reworking of the Matlab™ code but would ultimately prove to be a more general methodology. Each of the modifications described here would lead to greater accuracy in the predictions.

ACKNOWLEDGEMENTS

We gratefully acknowledge the support of the National Electric Energy Testing, Research, and Applications Center, NEETRAC, used for part of the work presented in this paper. We would also like to thank Mr. Bill Forrest, who pioneered the formulation of the problem from which the solution, presented in this paper, was derived.

REFERENCES

[1] Papoulis, A. and Pillai, S. U., "Probability, Random Variables and Stochastic Processes", Fourth Edition, McGraw-Hill, New York, 2002.

[2] Dodson, B., "Weibull Analysis", ASQ Quality Press, Milwaukee, 1994.

[3] Brown, R., "Electric Power Distribution Reliability", Marcel Dekker, New York, 2002.

[4] Billinton, R. and Allan, R., "Reliability Evaluation of Power Systems", Second Edition, Plenum Press, New York, 1996.

[5] Ibrahim, J., Chen, M.-H., and Sinha, D., "Bayesian Survival Analysis", Springer Verlag, NY, 2001.

[6] R. Billinton and W. Wenyuan Li, "Reliability Assessment of Electric Power Systems Using Monte Carlo Methods", New York: Plenum, 1994.

[7] W. Li, "Incorporating aging failures in power system reliability evaluation," IEEE Trans. Power Syst., Vol 17, p. 918, 2002.

[8] Forrest, I. W., "Predicting Medium Voltage Underground Power Cable Failures and Replacement Costs", presented at the Western Electric Power Institute Underground Distribution Workshop, Portland, OR, 1997.