New Probabilistic Method for Estimation of Equipment Failures and Development of Replacement Strategies

Miroslav Begovic[1], Petar Djuric[2], Joshua Perkel[1], Branislav Vidakovic[1], Damir Novosel[3]
[1]Georgia Institute of Technology, [2]Stony Brook University, [3]KEMA T&D Consulting

Abstract

When large amount of statistical information about power system component failure rate is available, statistical parametric models can be developed for predictive maintenance. Often times, only partial information is available: installation date and amount, as well as failure and replacement rates. By combining sufficiently large number of yearly populations of the components, 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.

1. Introduction

The problems of assessing the useful lifetime of equipment has been a focus of intense interest of electric utilities, especially in the circumstances when industry-wide restructuring and competition are tightening the operation and maintenance (O&M) budgets and managers are facing a dilemma of where to allocate the (often very limited) resources for the best possible use [3], [4]. An accurate model of power apparatus lifetime should contain a large number of factors, which are not practical for monitoring – a partial list should contain the initial quality and uniformity of the materials the equipment is made of (primarily the insulation), the history of exposure to moisture, impulse stress, mechanical stress, and many other factors. As those are neither available in typical situations (databases often do not even associate failures with the age), nor is their impact well documented and understood, the model that captures the essential behavior is, by necessity and for practical reasons, chosen to contain the most salient features known to be the strong determinants of lifetime. It is also assumed to be consistent with the Weibull distribution [2], [5], as it offers a degree of flexibility that other commonly used distributions do not.

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 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

$$P(t) = \int_0^t p(u)du. \qquad (1)$$

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

$$P_0(t) = 1-(1-P(t))^N \qquad (2)$$

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

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

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

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

0-7695-2507-5/06/$20.00 (C) 2006 IEEE

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 β>1 (β<1), the failure rate is increasing (decreasing) with time, while for β=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,

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

where u is a step function of time

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

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):

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

Arranged in a table, the number of failures per year per population would be as in Table 1.

Table 1. ***Cable population and the expected number of failures each year.***

#/yr	1	…	g	g+1	...	g+n+k
x_1	0		0	x_1a	...	$x_1(n+k)a$
x_2			0	0	...	$x_2(n+k-1)a$
x_3			0	0	...	$x_3(n+k-2)a$
x_4			0	0	...	$x_4(n+k-3)a$
. . .			. . .	. . .	...	
x_n			0	0	...	$x_n(k+1)a$
Sum			F_g	F_{g+1}	...	F_{g+n+k}

Total number of failures will therefore be:

$F_{g+1}=x_1a$

$F_{g+2}=(2x_1+x_2)a$

$F_{g+3}=(3x_1+2x_2+x_1)a$

$F_{g+4}=(4x_1+3x_2+2x_3+x_4)a$

$F_{g+n}=[nx_1+(n-1)x_2+(n-2)x_3+\ldots+x_n]a$

$F_{g+n+1}=[(n+1)x_1+nx_2+(n-1)x_3+\cdots+3x_{n-1}+2x_n]a$

$F_{g+n+2}=[(n+2)x_1+(n+1)x_2+nx_3+\cdots+4x_{n-1}+3x_n]a$

$F_{g+n+k}=[(n+k)x_1+(n+k-1)x_2+(n+k-2)x_3+\cdots(k+2)x_{n-1}+(k+1)x_n]a$

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

$$\Delta_i = F_i - f_i. \quad (25)$$

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

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

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 $\partial\Delta(a)/\partial a = 0$,

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

Solution to

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

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

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

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 = $1,2,...,n$ are numbers of installed miles from year 1 until year n. Let the vector $R = [r_1\ r_2\ \ r_n\ \ r_{n+k}]^T$

represent the quantities of cable removed from service in the years *1, 2, 3, …, n, …, n+k* respectively. In order to present the algorithm, we shall assume that the following is known: X_i , $i = 1,2,...,n$ – number of installed components from year *1* until year *n*; X_j , $j = n+1,n+2,...,n+k$ - number of installed components from year $n+1$ until year $n+k$; R_i , $i = 1,2,...,n$ - number of removed components from year *1* until year *n*.

Our objective is to find the numbers R_j , $j = n+1, n+2,..., n+k$ 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, ..., i, i*+1, ..., *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):

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

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 , $j = n+1,n+2,...,n+k$ represents the estimated number of failures in years *n+1, n+2, n+3, …, n+k,* respectively.

$$\begin{vmatrix} F_1 \\ F_2 \\ \\ F_{n-1} \\ F_n \end{vmatrix} = a \cdot \begin{vmatrix} 1^b & 0 & & & 0 \\ 2^b & 1^b & 0 & & 0 \\ & & & & \\ (n-1)^b & (n-2)^b & & 1^b & 0 \\ n^b & (n-1)^b & & 2^b & 1^b \end{vmatrix} \cdot X \tag{32}$$

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, ..., 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. Improved Formulation

Let there be M populations installed in years i_m, where $m = 1, 2, \ , M$, and $i_1 \le i_2 \le \ \le i_M$. Without loss of generality, we assume that $i_1 = 0$. In addition, let the *m*-th population be of size N_m. We also assume that the failure of a single component installed in year i_m is modeled by a parametric PDF, which is denoted by $p_m(t \mid \theta)$, where θ are the parameters to be estimated. If we know the PDF, we can compute the probability of failing of a particular component in the k -th year (where $k > i_m$) by

$$P_{m,k} = \int_{k-1}^{k} p_m(t \mid \theta)\, dt. \tag{33}$$

Let there be reports for the number of failures of components in years $j = 1, 2, \ , J,$ given by $n_{m,j}$ (number of failures from population *m* in the j^{th} year). We denote all the available failure data by $N_a = \{n_{1,1}, n_{1,2}, \ , n_{1,J}, n_{2,1}, \ , n_{2,J}, \ , n_{M,J}\}$. In general, we assume that we know the mathematical forms of $p_m(t \mid \theta)$ (note that these PDFs share the same vector of parameters θ, which is not known). Since we assumed that we know very little about θ, we model its prior as a constant.

There are two problems that we are interested in:

1. Given the functional forms $p_m(t \mid \theta)$ and past failure data, find the posterior PDF of θ, $p(\theta \mid N_a)$.
2. Using the obtained $p(\theta \mid N_a)$, predict future failures in the system.

Our approach is Bayesian, so we first find the posterior of θ, $p(\theta \mid N_a)$, in year J, and proceed by using it for prediction of failures of a given component in year $J+1$ according to

$$P_{m,J+1} = \int_J^{J+1} \int_\Theta f_m(t \mid \theta) p(\theta \mid N_a) d\theta dt \quad (34)$$

where $f_m(t \mid \theta)$ is the failure rate of the cables conditioned on θ,

$$f_m(t \mid \theta) = \frac{p_m(t \mid \theta)}{1 - \int_0^t p_m(\tau \mid \theta) d\tau} \quad (35)$$

and θ is the space of the unknown parameters.

First, suppose that we use the failures from population $m=1$ only. Then, for the probability of failures $n_{1,1}, n_{1,2}, \quad n_{1,J}$, we can write

$$P(n_{1,1}, n_{1,2}, \quad n_{1,J}) = \frac{N_1 !}{n_{1,r} ! \prod_{j=1}^{J} n_{1,j} !} P_{1,r}^{n_{1,r}} \prod_{j=1}^{J} P_{1,j}^{n_{1,j}} \quad (36)$$

where

$$n_{1,r} = N_1 - \sum_{j=1}^{J} n_{1,j}$$

$$P_{1,j} = \int_{j-1}^{j} p_1(t \mid \theta) d\theta \quad (37)$$

$$P_{1,r} = 1 - \sum_{j=1}^{J} P_{1,j} = \int_J^\infty p_1(t \mid \theta) dt.$$

Note that the probabilities $P_{1,j}$ carry information about θ, which we do not denote explicitly. Analogous expressions hold for the other populations. Since all the failures are considered independent, we can write for the joint probability mass function of the failures

$$P(n_{1,1}, n_{1,2}, \quad n_{1,J}, n_{2,1}, \quad , n_{M,J}) =$$

$$\prod_{m=1}^{M} \frac{N_m !}{n_{m,r} ! \prod_{j=1}^{J} n_{m,j} !} P_{m,r}^{n_{m,r}} \prod_{j=1}^{J} P_{m,j}^{n_{m,j}}. \quad (38)$$

Since our prior of θ is a constant, we write for the posterior

$$p(\theta \mid N_a) \infty P(n_{1,1}, n_{1,2}, \quad n_{1,M}, n_{2,1}, \quad , n_{M,J}). \quad (39)$$

From here on, we assume that the $p_m(t \mid \theta)$ is defined by the Weibull distribution, i.e.,

$$p_m(t \mid \theta) = \beta \alpha^{-\beta} (t - i_m)^{\beta - 1} e^{-\left(\frac{t - i_m}{\alpha}\right)^\beta}, \quad \alpha, \beta > 0 \quad (40)$$

where $\theta = (\alpha, \beta)$. It is readily shown that in this case

$$P_{m,j} = \int_j^{j+1} p_m(t \mid \theta) dt = e^{-\left(\frac{j - i_m}{\alpha}\right)^\beta} - e^{-\left(\frac{j + 1 - i_m}{\alpha}\right)^\beta} \quad (41)$$

$$P_{m,r} = e^{-\left(\frac{J - i_m}{\alpha}\right)^\beta} \quad (42)$$

and we deduce that

$$\log p(\theta \mid N_a) = const - \sum_{m=1}^{M} n_{m,r} \left(\frac{J + 1 - i_m}{\alpha}\right)^\beta + \sum_{m=1}^{M} \sum_{j > i_m}^{J} n_{m,j} \log\left(e^{-\left(\frac{j - 1 - i_m}{\alpha}\right)^\beta} - e^{-\left(\frac{j - i_m}{\alpha}\right)^\beta} \right) \quad (43)$$

With all these expressions, it is now relatively straightforward to compute the distribution of the posterior and the computation of the probabilities of failure. One method that could be applied is based on Monte Carlo computations.

4. Illustration of the Proposed Method

A synthesized data set is shown in Table 2. 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.

Table 2. *Modified data set used for verification of the algorithm.*

Year	Components installed	Components removed	Total Failures
1	22	10	0
2	44	15	2
3	63	16	0
4	82	10	1
5	104	9	10
6	193	18	6
7	215	16	9
8	329	28	9
9	370	61	13
10	417	60	10
11	453	43	17
12	510	29	25

13	437	17	28
14	395	19	32
15	61	4	48
16	17	3	30
17	16	1	43
18	0	0	45
19	0	0	57
20	0	0	53
21	0	0	68
22	0	0	75
23	0	0	67
24	0	0	81
25	0	0	88
26	0	0	100
27	0	0	84
28	0	0	127
29	0	0	154
30	0	0	139
31	0	0	137
32	0	0	156
33	0	0	151

The compliance of the data set with Weibull distribution is evidenced from Figure 1, which indicates a reasonably good fit.

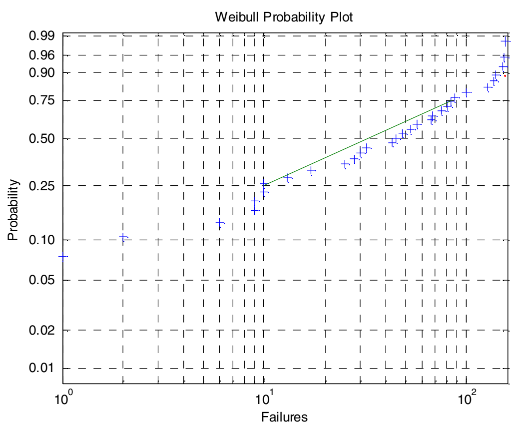

Figure 1. Weibull probability plot of failure data.

Table 3. ***Comparison between actual and estimated failures using the least squares parameter identification.***

Year	Actual Failures	Estimated Failures	Error (Act.-Est.)
1	0	0.024	-0.024
2	2	0.083	1.917
3	0	0.211	-0.211
4	1	0.453	0.547
5	10	0.831	9.169
6	6	1.396	4.604
7	9	2.257	6.743
8	9	3.476	5.524
9	13	4.900	8.101
10	10	6.921	3.079
11	17	9.754	7.246
12	25	13.489	11.511
13	28	18.076	9.924
14	32	23.318	8.682
15	48	29.114	18.886
16	30	35.148	-5.148
17	43	41.403	1.597
18	45	47.824	-2.824
19	57	54.371	2.629
20	53	61.031	-8.031
21	68	67.793	0.207
22	75	74.649	0.351
23	67	81.591	-14.591
24	81	88.613	-7.613
25	88	95.710	-7.710
26	100	102.880	-2.880
27	84	110.110	-26.110
28	127	117.410	9.590
29	154	124.770	29.230
30	139	132.190	6.810
31	137	139.660	-2.660
32	156	147.190	8.810
33	151	154.760	-3.760

The results from Table 2 are compiled into Figure 2, 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). Table 3 shows the comparison between estimated and observed number of failures (goodness of fit obtained by determining parameters of the Weibull distribution for the components).

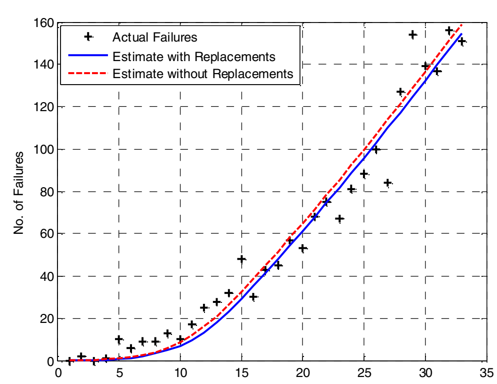

Figure 2. Estimated number of component failures assuming no replacements (upper curve) and replacements

(lower curve) superimposed over the actual numbers of component failures (shown as points labeled '+').

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:

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.

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 3 and 4 illustrate the confidence ranges (50% and 95%) obtained by running a Monte Carlo simulation on the same data set (described in Table 2).

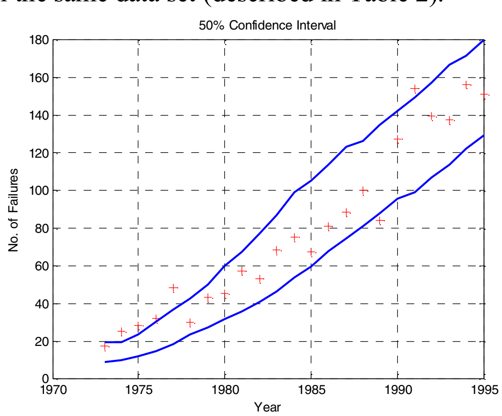

Figure 3. 50% confidence interval for estimated failures. Symbols '+' represent the actual observation data. The original data set is shown in Table 2.

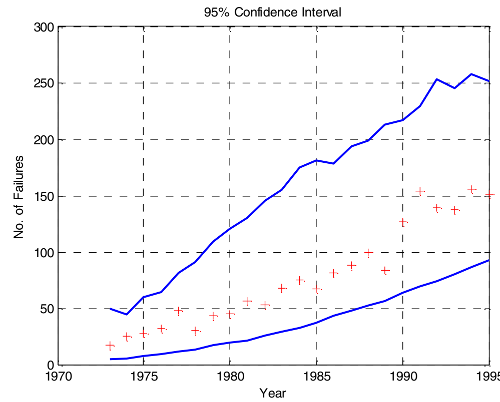

Figure 4. 95% confidence interval for estimated failures. Symbols '+' represent the actual observation data. The original data set is shown in Table 2.

5. Conclusions

The algorithm presented in this paper relies solely on basic chronological failure data to forecast the number of failures. As a consequence of the available data, some assumptions were made 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.

The algorithm can be used to forecast how actions in the present will impact the overall failure trend. Monte Carlo simulation [6] was employed to extract the confidence range of the estimated failures (or replacement rates needed to maintain a desired failure performance). By performing sufficiently large number of simulations (of the order of thousands), the algorithm yields a distribution for each of the parameters of interest and from these confidence intervals may be extracted.

The proposed 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.

6. Acknowledgment

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. Rick Hartlein for many inspiring and insightful conversations, and Mr. Bill Forrest, who pioneered formulation of the problem from which the solution, presented in this paper, was derived.

7. References

[1] Papoulis, A., 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., Allan, R., *Reliability Evaluation of Power Systems*, Second Edition, Plenum Press, New York, 1996.

[5] Ibrahim, J., Chen, M.-H., 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, pp. 918–923, Aug. 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, April 8, 1997.