EUROPEAN TRANSACTIONS ON ELECTRICAL POWER
Euro. Trans. Electr. Power 2009; **19**:631–641
Published online 12 March 2008 in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/etep.243

Evaluation aging failure probability of generating units using data analytic method

Seung-Hyuk Lee, In-Su Bae, Jin-O Kim*,† and Sung-Hoon Lee

Department of Electrical Engineering, Hanyang University, 17 Haengdang-dong, Sungdong-gu, Seoul, 133-791, Korea

SUMMARY

This paper presents a method to predict failure probability related to the aging of generating units in power systems. To calculate aging failure probability, Weibull distribution is usually used due to age-related reliability. The shape and scale parameters of Weibull distribution can be estimated by using hybrid method using both normal and Weibull distributions. In this paper, data analytic method applied to type II censoring is proposed which is relatively simpler and faster than the hybrid method for the estimation of parameters. The real historical data of combustion turbine generating units in Korean power systems are used in the case study to show how to calculate the mean life and its standard deviation by the proposed method. Copyright © 2008 John Wiley & Sons, Ltd.

KEY WORDS: aging failure; mean life; repairable failure; reliability; data analytic method; type II censoring

1. INTRODUCTION

Generally, repairable failure is defined as that it is repairable and can be operated well after repair, while aging failure as end-of-life failure that it is a part of wear-out and not able to be repaired any more [1]. Only the repairable failure of power system equipments have been considered to calculate mean life for the conventional methods, but it would be more elaborate if the non-repairable failure, namely, aging failures are considered. If certain equipments in a power system are aged, aging failures could become primary factors of system unreliability. In this case, consideration of repairable failures only may be likely to result in a misleading conclusion in system reliability [2]. Generally, generating units have relatively fewer aging failure data than any other equipment in power system because of their long life spans. Therefore, it is difficult to predict the mean life of generators due to the rarity of data compared to their long life span.

In order to evaluate the aging failure of equipments, it is important to obtain the mean life of equipment and its standard deviation as variables in power system [3]. There have been generally two ways to calculate the mean life and standard deviation of equipments, that is, the sample mean method (the average age method) and the method using probabilistic distribution [4].

The sample mean method is the most popular and simple one in statistical analysis and is to calculate the average age of dead equipments in historical data. However, it may be inappropriate for power system equipments since the survival probability is not involved for survivor equipments. The essential weakness of sample mean method is that it uses only information of retired equipments.

In the method using probabilistic distribution for the aging failure model of power system equipments, normal or Weibull distribution is often used. The simple procedure of maximum likelihood estimation (MLE) for normal distribution creates a sample mean which is still the average of ages of dead equipments. On the contrary, the MLE for a Weibull distribution leads to an extremely complex equation with multiple solutions for the shape and scale parameters, though it is more exact than normal distribution. Another method using estimated Weibull parameters by the mean and standard deviation of normal distribution, which is called the hybrid method in this paper, has been developed to calculate the mean life of equipments in power system [5]. The mean and standard deviation of normal distribution has been obtained using the least square error method, and the parameters of Weibull

*Correspondence to: Jin-O Kim, Department of Electrical Engineering, Hanyang University, 17 Haengdang-dong, Sungdong-gu, Seoul, 133-791, Korea.
†E-mail: jokim@hanyang.ac.kr

Copyright © 2008 John Wiley & Sons, Ltd.

distribution were estimated by using gradient descent method with the mean and standard deviation of normal distribution. However, this method still includes the complex calculating process.

This paper proposes data analytic method applied to type II censoring algorithm instead of the hybrid method to estimate the parameters of Weibull distribution for the assessment of the mean life and survival probability of equipments. The proposed approach has simple calculating process without deteriorating the accuracy. The case study in this paper was performed using the real historical data of combustion turbine generating units in Korean power systems, and the result from the conventional hybrid method is compared with those of proposed method.

2. APPROPRIATE DISTRIBUTION FUNCTION FOR A MEAN LIFE CALCULATION

The sample mean method is only based on data of retired equipments as mentioned above. However, the mean life of equipment group should be calculated considering both retired and surviving equipments. To calculate the mean life, Weibull distribution has the mathematical flexibility and is often used to obtain distribution on aging or aging failure process according to reliability theory [6]. The Weibull distribution is characterized by scale (α) and shape (β) parameters. The failure density function of the Weibull distribution is defined as

$$f(t) = \frac{\beta \cdot t^{\beta-1}}{\alpha^{\beta}} \exp\left[-\left(\frac{t}{\alpha}\right)^{\beta}\right] \tag{1}$$

where $t \geq 0$, $\beta > 0$, and $\alpha > 0$.

The survivor function is also expressed as

$$S(t) = \int_{r}^{\infty} f(t)\mathrm{d}t = \exp\left[-\left(\frac{t}{\alpha}\right)^{\beta}\right] \tag{2}$$

The Weibull distribution has no specific but wide variety of characteristics depending upon the values of the parameters (α, β) in its reliability functions. In other words, it can be shaped to represent many distributions. For this reason, the Weibull distribution has a very important role to play in the statistical analysis of experimental data. The Weibull distribution can be scaled and shaped by varying its shape parameters, β. $\beta < 1$ represents the debugging period, $\beta = 1$ the normal life period, and $\beta > 1$ the wear-out period. Aging failures are likely to happen during the wear-out period. So the Weibull distribution expresses the property of aging failure quite well.

3. A MEAN LIFE PREDICTION USING WEIBULL PARAMETERS ESTIMATION

In the preceding study [3], information of the retired and surviving equipments can be acquired by hybrid method using both normal and Weibull distribution functions. After obtaining the mean and standard deviation of the normal distribution as an initial estimate using least square error method, a set of nonlinear equations are set up and should be solved for the scale and shape parameters of the Weibull distribution. At this time, gradient descent method can be utilized for selecting the best estimates from the multiple solutions of the nonlinear equations [7]. This method tried to avoid a resolution process associated with a set of nonlinear simultaneous equations; however, the optimization approach to search the optimal solution still has a complexity with an iterative calculation process. Besides the methods mentioned above, the parameters of Weibull distribution could be estimated by using the MLE method. While it has an advantage that it can estimate each parameter of Weibull by estimating the maximum likelihood in case of existence of both censored and uncensored data, it may make a big error in case of rare or incomplete data like the failure data of power system equipment. Therefore, this paper proposes data analytic method applied to censoring algorithm for assessment of the mean life and failure probability of equipments, instead of hybrid method including the least square error method and gradient descent algorithm to estimate the parameters of Weibull distribution. Data analytic method has advantages that it does not require a resolution process for gradient descent method nor iterative calculation process for the optimization approach.

Figure 1 shows the procedure to calculate the mean life of equipments with the flowchart of these two methods.

Generally, the data of equipments in power system have incompleteness characteristics due to the scarcity of their aging failure data in contrast to their relatively long life. Thus, there needs a stochastic approach known as data analytic method for this

Copyright © 2008 John Wiley & Sons, Ltd.

DOI: 10.1002/etep

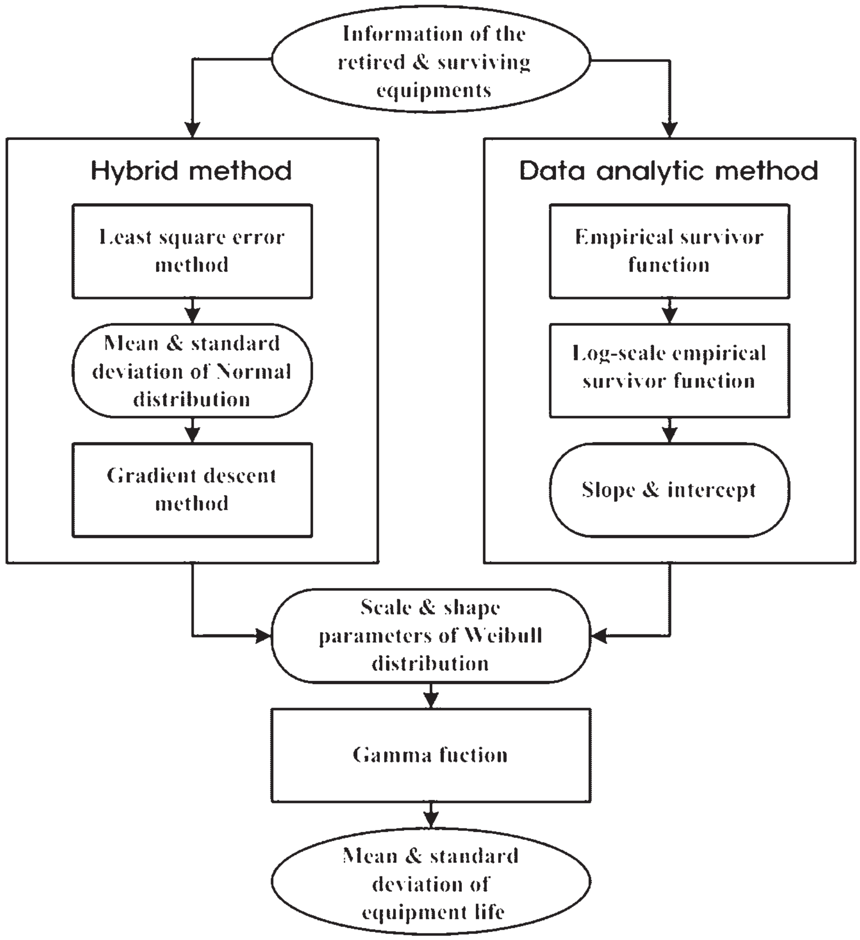

Figure 1. Flowchart of mean life calculation procedure.

incomplete data analysis. Data analytic method can be divided into two censoring methods according to the acquired data: type I and II censorings [8,9].

3.1. *Type I censoring (left-censored observation)*

The aging failures $t_i (i = 1, 2, ..., n)$ of all equipments are known only to have retired prior to some observation time, where t_i is the lifetime of ith equipment for total n equipments.

3.2. *Type II censoring (right-censored observation)*

Unlike type I censoring, both aging failure and survivor equipments exist for the observation time, and the lifetime of survivor ones are known to exceed a certain time. Such an observation is commonly made in power system and called as right-censored observation. If the statistical data for calculating the mean life of equipments are right-censored, then survivor probabilities should be considered for the survivor equipments.

The data analytic method applied to type II censoring is utilized for construction of the empirical survivor function $\hat{S}(t)$, which can be used afterwards to estimate the parameters of Weibull distribution. If the observed lifetimes are put in ascending order, then

Copyright © 2008 John Wiley & Sons, Ltd.

DOI: 10.1002/etep

the empirical survivor function [8] is defined as

$$\hat{S}(t_i) = 1 - \frac{(i-1)}{n}, \quad t_1 < t_2 < \cdots < t_n \tag{3}$$

The empirical survivor function evaluated at a point just greater than a certain time t_i is also expressed as

$$\hat{S}(t_i + 0) = 1 - \frac{(i-1)}{n} - \frac{1}{n} = 1 - \frac{i}{n} \tag{4}$$

Thus, the average value of $\hat{S}$ evaluated near t_i can be obtained as

$$\frac{1}{2}\{\hat{S}(t_i) + \hat{S}(t_i + 0)\} = 1 - \frac{(i-1)}{n} - \frac{1}{2n} = 1 - \frac{(i-1/2)}{n} \tag{5}$$

A graphical description [5] of the empirical survivor function is to plot the point represented as Equation (6). In Figure 2, the above plots show the empirical survivor functions of certain equipments, and the lower ones illustrate the log-scaled estimate of $S(t)$ and approximate certain confidence bands for $S(t)$.

$$(x_i, y_i) = \left(t_i,\ 1 - \frac{(i-1/2)}{n}\right) \tag{6}$$

where x_i is a set of t_i; y_i is a set of $1 - (i - 1/2)/n$.

For the Weibull distribution, mathematical relationship such as Equation (7) is given from Equation (2).

$$\ln\{-\ln S(t)\} = \beta \ln t - \beta \ln \alpha \tag{7}$$

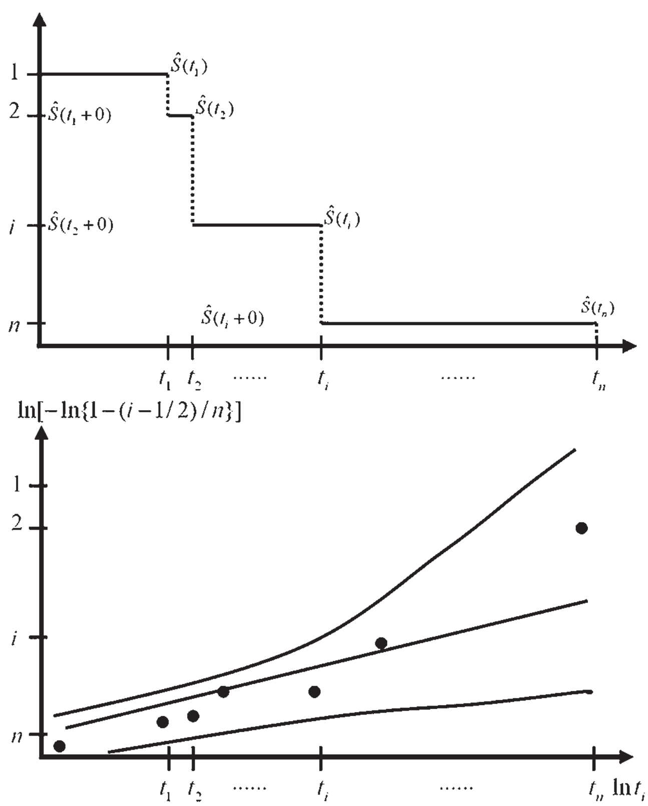

Figure 2. Plots of $\hat{S}(t_i)$ using data analytic method.

Copyright © 2008 John Wiley & Sons, Ltd.

DOI: 10.1002/etep

Consequently, it can be seen that a plot of $\ln[-\ln\{1-(i-1/2)/n\}]$ against $\ln t_i$ should be linear. If the slope and intercept of the linear regression plot are a and b, then estimates of shape and scale parameters, β and α, become a and $\exp(-b/a)$, respectively [6].

Finally, the mean life (μ) and its standard deviation (σ) are to be calculated using the shape and scale parameters obtained from the hybrid method or the proposed data analytic method using type II censoring, and the calculation formulas are given as Equations (8) and (9) [4,10].

$$\mu = \alpha \cdot \Gamma\left(1+\frac{1}{\beta}\right) \tag{8}$$

$$\sigma^2 = \alpha^2\left[\Gamma\left(1+\frac{2}{\beta}\right) - \Gamma^2\left(1+\frac{1}{\beta}\right)\right] \tag{9}$$

where the gamma function $\Gamma(*)$can be approximated [4] as

$$\Gamma(y) = \sqrt{2\pi}y^{(y-0.5)}\mathrm{e}^{-y}\left(1+\frac{1}{12y}\right) \tag{10}$$

4. FAILURE PROBABILITY PREDICTION USING CONDITIONAL PROBABILITY

For the precise prediction of failure probability, the conditional probability can be applied to cumulative failure distribution (CFD) function of Weibull [9]. The CFD function of Weibull is the probability of failure up to time t and is expressed as

$$Q(t_i) = \left(1-\exp\left(-\frac{1}{\alpha_i}(t_\phi - t_{0,i})\right)^{\beta_i}\right) \tag{11}$$

where $\alpha_i > 0, \beta_i > 0, (t-t_i) \geq 0, 0 \leq Q_i(t) \leq 1$; t_ϕ is the observation time; $t_{0,i}$ is the installation time of equipment I; $t_i = (t_\phi - t_{0,i})$ is the age of equipment i.

If we consider the equipments that have survived into the present, the future failure can be predicted using conditional probability. To predict the aging failure probability, Equation (11) should be modified to include the condition that the predictions are beyond the present time (t_ϕ); that is, the age, t_i of Equation (11), can be treated as future time.

The prediction of aging failure is expressed as a probability that the equipments would be retired between the present and a certain time in the future under the condition that it survives until the present, and the conditional probability is defined as Equation (12). Figure 3 depicts the intersection of two time interval sets, which shows the definition of the conditional probability as appeared in

$$Q(t_i, t_\phi) = Q(t_\phi \leq t \leq t_i \middle| t \geq t_\phi) \tag{12}$$

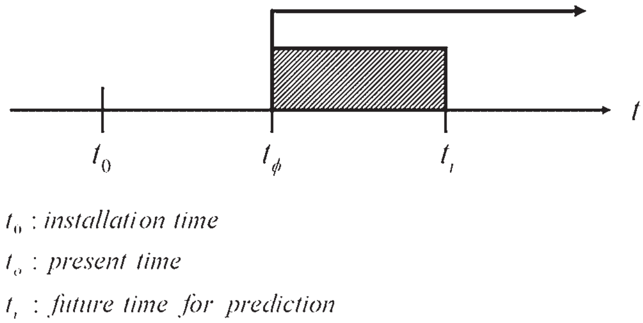

Figure 3. The relation of time for conditional probability.

Equation (12), which is expressed as the passage of time, can be transformed as

$$Q(t_i,t_\phi) = \frac{Q(t_i) - Q(t_\phi)}{1 - Q(t_\phi)} \tag{13}$$

5. CASE STUDY

This paper applied the data analytic method to the real historical data of combustion turbine generating units in Korean power systems, and compared the estimated results with the conventional hybrid method. Thirty-six generators are observed for the case study, and among these generators, 16 generators have been failed, retired, and replaced due to aging failures during the past 41 years. The assumptions used for the case study are as follows:

- Combustion turbine generating units are used without maintenance and inspection.
- The environment, operation methods, material, manufacturer, and type of all combustion turbine generating units are identical and the characteristics are unchanged within observed period.

Table I. Raw data of combustion turbine generators.

Gen. ID	Installed year	Retired year
1	1962	1974
2	1962	1975
3	1967	1974
4	1962	1989
5	1962	1993
6	1962	1968
7	1968	1993
8	1977	1989
9	1968	1993
10	1975	1993
11	1977	1993
12	1977	1993
13	1977	1997
14	1979	1997
15	1977	1997
16	1979	1997
17	1992	
18	1992	
19	1996	
20	1997	
21	1992	
22	1993	
23	1992	
24	1993	
25	1994	
26	1997	
27	1992	
28	1994	
29	1993	
30	1993	
31	1993	
32	1993	
33	1995	
34	1996	
35	2000	
36	2001	

Copyright © 2008 John Wiley & Sons, Ltd.

The factors of aging failures contain not failures with a fire or an explosion but inappropriate operation or component defects from its production stage, etc. They are usually related to the causes to reduce the lifetime of equipment. It is assumed that all of the installation, operation, and maintenance conditions are identical for the observed generators, and the reference year is 2003.

Table I shows the original data of the observed combustion turbine generators. For the retired generators, their ages are the difference between the retired and the installed year, while for the surviving generators, the ages are the years up to the present from the installed year. In this table, the age of number 6 generator is relatively short and it seems the case that this generator is eliminated after retiring without any maintenance due to any circumstances of generating company. The numbers of exposed and retired generators for each year of age in Table I can be counted and assorted in Table II in the ascending order of their ages. Using the values appeared in Table II, the hybrid method is applied to estimate the shape and scale parameters of Weibull distribution. The

Table II. Numbers of exposed and retired combustion turbine generators.

Age (year)	No. of exposed generators	No. of retired generators
0	36	
1	36	
2	36	
3	35	
4	34	
5	34	
6	34	1
7	32	1
8	30	
9	29	
10	27	
11	21	
12	16	2
13	16	1
14	16	
15	16	
16	16	2
17	16	
18	16	3
19	16	
20	16	2
21	16	
22	16	
23	16	
24	16	2
25	14	
26	14	1
27	9	
28	9	
29	8	
30	8	
31	8	1
32	8	
33	8	
34	8	
35	8	
36	6	
37	5	
38	5	
39	5	
40	5	
41	5	

Copyright © 2008 John Wiley & Sons, Ltd.

DOI: 10.1002/etep

Table III. Data rearranged for data analytic method: type II censoring.

No. of retired generator (i)	Age (year) (T_i)	Combustion turbine generator ID number
1	6	6
2	7	3
3	12	1
4	12	8
5	13	2
6	16	11
7	16	12
8	18	10
9	18	14
10	18	16
11	20	13
12	20	15
13	25	7
14	25	8
15	27	4
16	31	5

resulted values are 19.541 and 2.593, and the mean life and its standard deviation for the observed generators are calculated as 17.341 and 7.206, respectively, using Equations (8) and (9).

Table II is rearranged into Table III in order to conduct the proposed method using data analytic method applied to type II censoring. The retired generators are tabulated in chronological order with their ages and ID numbers.

The 16 retired generators are mapped into Figure 4, which shows the plots of $\ln[-\ln\{1-(i-1/2)n\}]$ against $\ln t_i$ using Equation (6) by data analytic method applied to type II censoring algorithm. In this expression, n denotes the total number of retried generators (in this case, $n = 16$). As can be seen in Figure 4, all points are regressed on a straight line and it creates the empirical survival function $\hat{S}(t)$ in log scale. The fitted straight line in Figure 4 yields slope and intercept to 2.680 and - 8.038 and therefore, estimates for β and α become 2.680 and $\exp(8.038/2.680) = 20.070$.

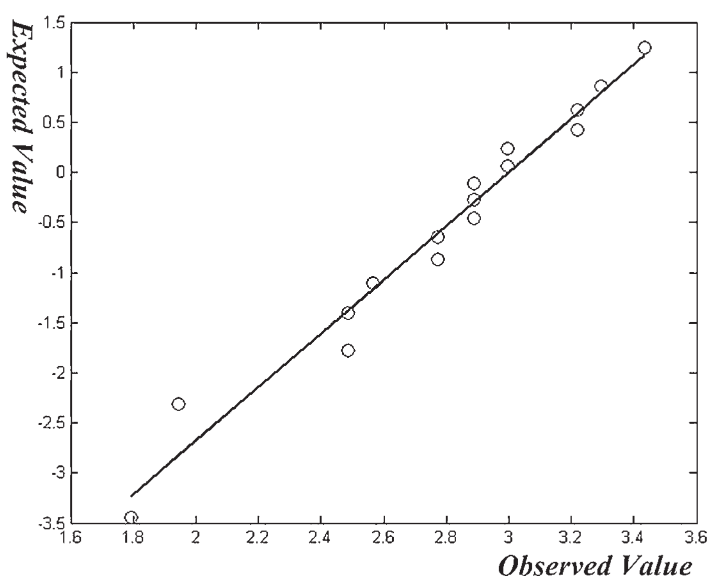

Figure 4. Linear regression of empirical survival function.

Copyright © 2008 John Wiley & Sons, Ltd.

DOI: 10.1002/etep

Table IV. Results of parameter estimation.

	Using hybrid method	Using MLE method	Relative error (%)	Using data analytic method	Relative error (%)
Shape parameter	2.593	2.897	11.724	2.680	3.355
Scale parameter	19.541	19.931	1.996	20.070	2.707
Mean	17.341	17.757	2.399	17.875	3.079
Standard deviation	7.206	7.195	0.153	7.193	0.180

Table IV shows that the estimates of the shape and scale parameters for the Weibull distribution using the hybrid method, MLE method, and data analytic method. However, as mentioned before, MLE method show the larger errors in shape parameter than the data analytic method proposed in this paper, in case of small number of data or incompleteness such as the failure data of generating units as shown in Figure 5 and Table IV.

The estimates of the mean life and its standard deviation for the combustion turbine generators are also given in Table IV. In this table, it can be seen that the results from these two methods are quite close within tolerance of 5%, but data analytic method is simpler than the hybrid method in the calculation process, since the hybrid method combines the results of normal and Weibull distributions where gradient descent method should be applied for avoiding the resolution process. Furthermore, as mentioned previously, data analytic method is useful for the data with incomplete and rare characteristics. As can be seen in Table III, the number of paired data which should be mapped into the statistical space is only the one of retired generators, and therefore, the linear regression only for them is quite simple compared with the hybrid method using both normal and Weibull distribution functions.

The CFD is obtained by substituting the estimated parameters in Table IV, and is depicted in Figure 6, which provides the fact that the mean life of combustion turbine generator is between 17 and 18 years (about 50% of $Q(t)$) and after 40 years, the aging failure probability would be almost 100%. Using this CFD, the probability of aging failure in the future can be predicted. For example, when the present time is at 10 years, the aging failure probability of combustion turbine generator that would be 30 years can be calculated as 93.82% as expressed in

$$Q(30, 10) = \frac{0.947 - 0.143}{1 - 0.143} = 0.9382 \tag{14}$$

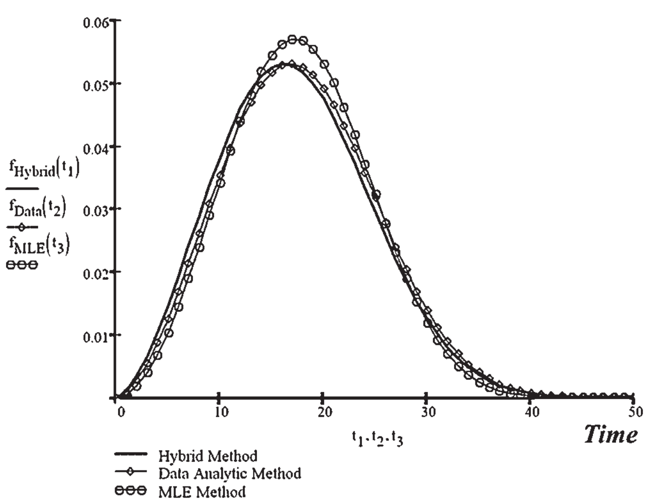

Figure 5. Weibull distribution function according to each method.

Copyright © 2008 John Wiley & Sons, Ltd.

DOI: 10.1002/etep

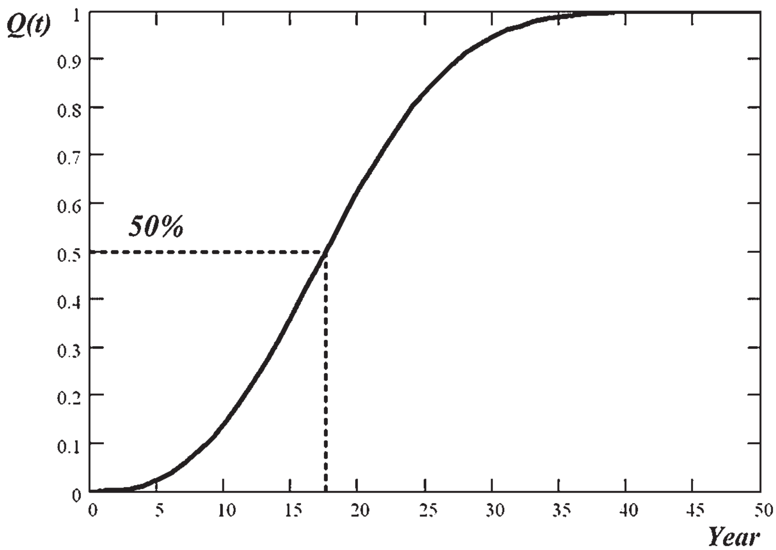

Figure 6. CFD of Weibull for combustion turbine generators.

6. CONCLUSIONS

This paper proposes the estimation of the Weibull parameters using data analytic method and presents the method to predict the aging failure probability considering the conditional probability. The proposed procedure for mean life estimation is based on all the information in equipment including both retired and surviving components. In the case study, it is clear that the proposed method provides not only the comparability in accuracy but also more simple calculation process than the hybrid method. For the precise prediction of aging failure probability, the conditional probability is also applied to CFD function of Weibull distribution. Furthermore, using the proposed method, reliability evaluation related to aging failure can be assessed easily for the power system equipments such as transformer, reactors, cables and generators.

7. LIST OF SYMBOLS AND ABBREVIATIONS

7.1. Symbols

a	slope of the linear regression plot
b	intercept of the linear regression plot
$f(t)$	failure density function
n	number of equipments
Q	cumulative failure distribution function
$S(t)$	survivor function
$\hat{S}(t)$	empirical survivor function
$t_{0,i}$	installation time of equipment i
t_i	life time of equipment i
t_ϕ	observation time
μ	mean life of generating unit
x_i	set of t_i
y_i	set of $1-(i-1/2)/n$
α	scale parameter of Weibull distribution
β	shape parameter of Weibull distribution

Copyright © 2008 John Wiley & Sons, Ltd.

σ standard deviation of generating unit's life
$\Gamma(*)$ gamma function

7.2. Abbreviations

CFD cumulative failure distribution
MLE maximum likelihood estimation

REFERENCES

1. Schilling MT, Praca JCG, de Queiroz JF, Singh C, Ascher H. Detection of ageing in the reliability analysis of thermal generators. *IEEE Transactions on Power Systems* 1988; **3**(2):490–499.
2. Beehler ME. Reliability centered maintenance for transmission systems. *IEEE Transaction on Power Delivery* 1997; **12**(2):1023–1028.
3. Cho SD. Three-phase transformer model and parameter estimation for ATP. *Journal of Electrical Engineering & Technology* 2006; **1**(3):302–307.
4. Li W. Incorporating aging failures in power system reliability evaluation. *IEEE Transaction on Power Systems* 2002; **17**(3):918–923.
5. Li W. Evaluating mean life of power system equipment with limited end-of-life failure data. *IEEE Transaction on Power System* 2004; **19**(1):236–242.
6. Billinton R, Allan RN. *Reliability Evaluation of Engineering Systems*. Plenum Press: New York, 1992.
7. Rao SS. *Engineering Optimization*. Wiley-Interscience: New York, 1995.
8. Crowder MJ, Kimber AC, Smith RL, Sweeting TJ. *Statistical Analysis of Reliability Data*. Chapman and Hall: London, 1991.
9. William QM, Luis AE. *Statistical Methods for Reliability Data*. Wiley: New York, 1998.
10. Bucci RM, Rebbapragada RV, McElroy AJ, Chebli EA, Driller S. Failure prediction of underground distribution feeder cables. *IEEE Transaction on Power Delivery* 1994; **9**(4):1943–1955.

Copyright © 2008 John Wiley & Sons, Ltd.

DOI: 10.1002/etep