International Journal of Mechanical, Industrial and Aerospace Sciences
ISSN: 2517-9950
Vol:9, No:6, 2015

Estimating the Life-Distribution Parameters of Weibull-Life PV Systems Utilizing Non-Parametric Analysis

Saleem Z. Ramadan

***Abstract*—**In this paper, a model is proposed to determine the life distribution parameters of the useful life region for the PV system utilizing a combination of non-parametric and linear regression analysis for the failure data of these systems. Results showed that this method is dependable for analyzing failure time data for such reliable systems when the data is scarce.

***Keywords*—**Masking, Bathtub model, reliability, non-parametric analysis, useful life.

I. Introduction

IN determining the feasibility of owning a PV system, the total cost of owner ship (TCO) for the system must be calculated. This cost includes (but not limited to) the initial purchasing cost of the system, the operating costs of the system, and the downtime cost of the system [1]. While the initial cost does not depend on the life distribution of the PV system, the downtime cost and parts of the operating cost do. Maintenance cost (corrective, preventive, and predictive) is a major part of the operating cost, which depends heavily on the life distribution of the PV system as the life distribution of the PV system governs its failure pattern. Another maintenance-related cost is the downtime cost, which includes the costs related to the stoppage of the PV system (loss of potential revenue) while performing the maintenance. Energy consumption is another part of operating costs, but this part does not apply in the case of PV systems as PV systems consume free fuel (solar radiation).

In evaluating the payback period or the energy price per KWh of the PV system, the useful life of the system must be known. Moreover, to evaluate the maintenance cost and the downtime cost for the PV system (which will be needed to calculate the payback period or the energy price per KWh), the life distribution of the useful life region for the system must be known. Assuming that the PV system works without interruption will underestimate the TCO of the PV system. This will affect the payback period and the energy price per KWh calculations which may lead to a wrong decision about the feasibility of the PV system.

Typically, the useful lives of the PV systems are in the range of 15- 20 years with the inverter failure is the main cause for the system failure [2], [3]. Unfortunately, there is a shortage in failure data sets for PV systems as these systems are relatively new. Moreover, the available data sets limited to time between failures without much information about the reasons of the failure.

This lack of information about the reasons of the failure (known in literature as masking problem) makes the extraction of the useful life region limits hard to obtain as the failure reasons are mixed between wear out, random, and manufacturing defects [4]-[7]. In this paper, a model for determining the useful life region limits and the life distribution governing this region for Weibull life PV systems is proposed utilizing non-parametric and linear regression analysis for the failure time data points. The effectiveness of the model is illustrated using simulated data.

The rest of the paper will be organized as follows: Section II will present the derivation of the model, Section III will utilize simulated data for discussion, and Section IV will conclude.

II. Derivation of the Proposed Model

For Weibull-life systems, the infant mortality region of the bathtub model is characterized by decreasing hazard rate function because the reliability of the system increases as the time goes on, while the opposite happens in the aging region as the reliability of the system decreases. Unlike these two regions, in the useful life region of the bathtub model, the hazard rate stays constant [8]. This means that the slope of the hazard rate function in the useful region is constant. This characteristic of the useful life region can be exploited in determining the limits of this region.

The slope of the hazard rate function in the useful region of the bathtub model can be obtained from the slope of the best line fit in this region. The best line fit can be found by simple linear regression model between the failure time data points and the corresponding hazard rate values at each of the failure time data points in the useful life region.

Hazard rate is defined as the conditional probability of failure in the interval t to $(t+dt)$, given that there was no failure at t. Mathematically, it can be calculated as [9]:

$$h(t_i) = \frac{f(t_i)}{R(t_i)} \tag{1}$$

The non-parametric formula for the hazard rate function can be derived from (1) as:

R. Z. Saleemis with Mechanical and Industrial Engineering Department, Applied Science Private University, Amman-Jordan-11931, P.O. Box 166(e-mail: s_ramadan@asu.edu.jo).

International Journal of Mechanical, Industrial and Aerospace Sciences
ISSN: 2517-9950
Vol:9, No:6, 2015

$$f(t_i)=\frac{df(t_i)}{dt_i}\cong\frac{\Delta F(t_i)}{\Delta t_i}=\frac{F(t_{i+1})-F(t_i)}{t_{i+1}-t_i}, \qquad (2)$$

where

$$F(t_i)=\frac{i}{N} , \qquad (3)$$

and i is the index for the failure time data point under consideration, and N is the total number of data points. Substituting (3) into (2) gives a non-parametric equation for the life distribution:

$$f(t_i)=\frac{\frac{i+1}{N}-\frac{i}{N}}{t_{i+1}-t_i}=\frac{1}{N(t_{i+1}-t_i)} \qquad (4)$$

The reliability distribution (complementary of failure distribution) also can be expressed in a non-parametric form utilizing (3) as:

$$R(t_i)=1-F(t_i)\cong 1-\frac{i}{N}=\frac{N-1}{N} \qquad (5)$$

Substituting (4) and (5) into (1) gives a non-parametric form for the hazard rate function as:

$$h(t_i)=\frac{\frac{1}{N(t_{i+1}-t_i)}}{\frac{N-i}{N}}=\frac{1}{(t_{i+1}-t_i)(N-i)} \qquad (6)$$

Equation (6) was reported by many other authors like [8].

Consider a set of ordered ungrouped complete failure data point t as:

$$t=\{t_1,t_2,..,t_{\tau 1},t_{\tau 1+1},t_{\tau 1+2},...,t_{\tau 2},t_{\tau 2+1},t_{\tau 2+2},...,t_N\},$$

where t_1 is the index of the failure time data point representing the start of the useful life, t_2 is the index of the failure time data point representing the end of the useful life. Define $\dot{t}$ as the set of failure time data points without the last failure time data point as:

$$\dot{t}=t-t_N,$$

Moreover, let h be the set of hazard rate values corresponding to the failure time data set $\dot{t}$ as:

$$h=\{h(t_1),h(t_2),...,h(t_{\tau 1}),...,h(t_{\tau 2}),h(t_{\tau 2+1}),...,h(t_{N-1})\}$$

Given the failure time data points (t_i) and the corresponding hazard rate

$$\{h(t_i),t_i : i=t_1,t_2,..,t_{\tau 1},t_{\tau 1+1},t_{\tau 1+2},...,t_{\tau 2}\}$$

The least square estimates to the slope m of the regression line between $h(t_i)$ and t_i as a function of τ_1 and τ_2 is computed as:

$$\hat{m}(\tau_1,\tau_2)=\frac{\sum_{i=\tau_1}^{\tau_2}\dot{t}\left(h(t_i)-\bar{h}\right)\left(h(t_i)-\bar{h}\right)}{\left(t_i-\bar{t}\right)^2}, \qquad (7)$$

where $\hat{m}$ is the least square estimate of the regression line, $\bar{t}$ is the average of the failure time data point set $\dot{t}$, and $\bar{h}$ is the average of the hazard rate set h .

In this paper, the time between failures for the PV systems is assumed to follow Weibull distribution. According to [8], the shape parameter of the Weibull distribution can be estimated using:

$$\hat{\theta}=\frac{\sum_{i=\tau_1}^{\tau_2}t_i^{\hat{\beta}}\ln t_i+(n-r)t_{\tau 2}^{\hat{\beta}}\ln t_{\tau 2}}{\sum_{i=\tau_1}^{\tau_2}t_i^{\hat{\beta}}+(n-r)t_{\tau 2}^{\hat{\beta}}}-\frac{1}{\hat{\beta}}=\frac{1}{r}\sum_{i=\tau_1}^{\tau_2}\ln t_i , \qquad (8)$$

where $r = card\ \{\tau_1, \tau_1 + 1,..., \tau_2\}$ and card(.) is the cardinality of the set.

Equation (8) can be solved numerically for β using Newton-Raphson procedure to find an estimation for β, $\hat{\beta}(\tau_1,\tau_2)$, as a function of τ_1 and τ_2 .

Fig. 1 shows a typical bathtub model. It is clear from Fig. 1 that as we move from the left toward τ_1 on the time axis, the slope of the hazard rate function increases until it reaches zero at τ_1. Moreover, as we move away from τ_2 to the left, the slope of the hazard rate function increases. This means that $\hat{\beta}\approx 1$ and $\hat{m}\approx 0$ between τ_1 and τ_2 and in their vicinities.

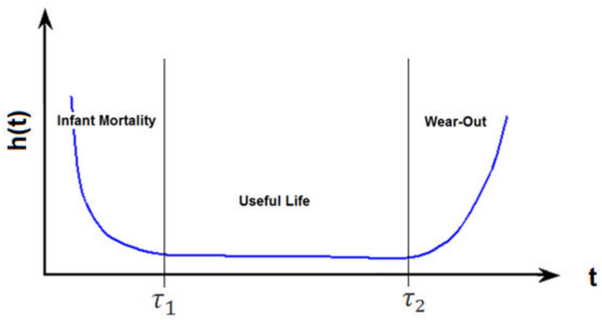

Fig. 1 Typical bathtub model

Exploiting that $\hat{\beta}\approx 1$ and $\hat{m}\approx 0$ between τ_1 and τ_2 and in their vicinities, we can build our objective function as:

$$Z = abs\left[\hat{m}(\tau_1,\tau_2)-\hat{\beta}(\tau_1,\tau_2)+1\right] \qquad (9)$$

Equation (9) suggests that in the useful life region for a Weibull-life PV systems Z must equal to zero. This is true because at that region the slope of the hazard rate function is zero and the shape parameter for the life distributing is one thus Z must be zero. Because the Z is calculated using sample of the failure time data, we should expect that Z will not have exactly zero, but something around this value. The value of zero will not be achieved unless there is a perfect data which is hard to obtain in real life situations.

The failure time data between τ_1 and τ_2 can be used to estimate the characteristic life of the PV systems as;

$$\hat{\theta} = \frac{1}{r}\sum_{i=\tau_1}^{\tau_2} \ln t_i \tag{10}$$

As a matter of fact, (9) can be used as our objective function for the proposed model, which can be used to find τ_1 and τ_2. Finding the optimal τ_1 and τ_2 determines the useful life region of the PV system.

Based on (9), the proposed model is as

$$\begin{aligned}
&\min Z = abs\left[\hat{m}(\tau_1,\tau_2) - \hat{\beta}(\tau_1,\tau_2) + 1\right] \\
&\tau_1 \geq 1 \\
&\tau_1 \leq \tau_2 \leq N-1 \\
&\tau_1, \tau_2 \geq 0, \text{integer}
\end{aligned}$$

where $\hat{m}$ and $\hat{\beta}$ are the slope of the regression line and the shape of the Weibull distribution parameter in the useful life region between τ_1 and τ_2.

Full enumeration technique will be employed to solve this model. The number of combinations need to be evaluated is $N_c = \sum_{i=1}^{N-1} i$. As the failure time data usually scarce for the PV systems, the number of combinations need to be evaluated is within the capabilities of most modern computers. For example if we have 1000 failure time data points, then the number of combinations need to be evaluated is $\sum_{i=1}^{999} i = 499500$ combinations, which is reasonable for most of the modern computers.

The procedure followed in this paper to find the optimal values of τ_1 and τ_2 is as:

Start

Step 1

Enumerate the combinations for the failure time data points and give each combination a serial number starting from 1: $c = \{c_1, c_2, ..., c_{N_c}\}$.

Set k=0

Step 2

Do while k < N_c

1. k=k+1
2. Set $D = \{t_i : i \in [c_{k1}, c_{k2}]\}$, where c_{k1} and c_{k2} are the first value and the second value in c_k, respectively.
3. Calculate the hazard rate h with equation (6) using the failure time data D.
4. Calculate the slope of the regression line $\hat{m}$ with equation (7) using the values of D and h.
5. Calculate $\hat{\beta}$ using equation (8) where $r = card(D)$
6. Calculate Z with equation (9) using the values of $\hat{m}$ and $\hat{\beta}$.
7. Store the value of Z along with its corresponding k value

End

Step 3

Choose the minimum value of Z as the optimal value and retrieve the corresponding optimal combination τ_1 and τ_2 using the corresponding k value.

Step 4

Use τ_1 and τ_2 and the corresponding D set to find the characteristic life $\hat{\theta}$ with (10) and retrieve the value for the shape parameter $\hat{\beta}$.

End

The above non-parametric procedure can be used to determine the useful life region of the Weibull-life PV systems and determine the life distribution parameters for the useful life region.

III. Illustrative Examples

Four illustrative examples using simulated data will be used to illustrate the effectiveness of this model. The first example will simulate data only in the useful life region. The second example will use simulated data from infant mortality and useful life regions. The third example will use simulated data from useful life and aging regions. And finally, the fourth example will use a simulated data from all three regions.

A. Example 1

Table I shows the pertaining failure time data for example 1. The failure time data (years) is simulated from a Weibull distribution with $\beta = 1$ and $\theta = 1.6$.

TABLE I
Failure Time Data for Example 1

2.0695	3.285	0.0966	0.006	2.0807
0.6825	0.1272	2.0023	1.1071	1.5107
6.825	0.3807	0.8319	1.8794	0.869
2.7308	0.7918	0.9206	0.7167	0.3098
5.5971	0.1143	2.6753	1.4214	1.6911
0.3631	0.3518	0.1397	0.4155	2.7609
0.1321	1.2889	1.3137	1.2121	0.1705
3.1936	5.2825	0.9347	0.9584	1.4796
1.5117	1.2859	0.8188	3.225	1.3022
4.778	1.8619	0.1939	0.5209	1.7917

Applying the proposed model, the optimal solution was $\tau_1 = 1$ and $\tau_2 = 50$ with $\hat{\beta} = 1.09$ and $\hat{\theta} = 1.56$. It is clear that the model was able to predict that there is only one region and the whole data set came from the useful life region that starts from 0.006 years and ends at 6.825 years.

B. Example2

Table II shows the pertaining failure time data for example 2. The data for this example is the same data used in example 1 added to it 10 failure data points from infinite mortality region from a Weibull distribution with $\beta = 0.86$ and $\theta = 0.32$. It should be clear from the data that the infant mortality data points overlap with useful life data to simulate the real life situation.

TABLE II
FAILURE TIME DATA FOR EXAMPLE 2

0.0039	2.0695	3.285	0.0966	0.006	2.0807
0.0999	0.6825	0.1272	2.0023	1.1071	1.5107
0.5016	6.825	0.3807	0.8319	1.8794	0.869
0.1743	2.7308	0.7918	0.9206	0.7167	0.3098
0.175	5.5971	0.1143	2.6753	1.4214	1.6911
0.1047	0.3631	0.3518	0.1397	0.4155	2.7609
0.1009	0.1321	1.2889	1.3137	1.2121	0.1705
0.6425	3.1936	5.2825	0.9347	0.9584	1.4796
1.407	1.5117	1.2859	0.8188	3.225	1.3022
0.2395	4.778	1.8619	0.1939	0.5209	1.7917

Applying the proposed model, the optimal solution was $\tau_1 = 3$ and $\tau_2 = 60$ with $\hat{\beta} = 0.99$ and $\hat{\theta} = 1.4$. The model was able to detect the two regions, the infant mortality region and the useful life region. Note that the model predicted that there is no third region, i.e., aging region because $\tau_2 = 60$. The model detected the beginning of the useful life region to be 0.0966 years which is close to the actual value of 0.006.

C. Example 3

Table III shows the pertaining failure time data for example 3. The data for this example is the same data used in example 1 added to it 10 failure data points from aging region from a Weibull distribution with $\beta = 83$ and $\theta = 6.56$. It should be clear from the data that the aging data points overlap with useful life data to simulate the real life situation.

TABLE III
FAILURE TIME DATA FOR EXAMPLE 3

2.0695	3.2850	0.0966	0.0060	2.0807	6.4353
0.6825	0.1272	2.0023	1.1071	1.5107	6.3657
6.8250	0.3807	0.8319	1.8794	0.8690	6.6070
2.7308	0.7918	0.9206	0.7167	0.3098	6.4703
5.5971	0.1143	2.6753	1.4214	1.6911	6.6311
0.3631	0.3518	0.1397	0.4155	2.7609	6.5353
0.1321	1.2889	1.3137	1.2121	0.1705	6.3927
3.1936	5.2825	0.9347	0.9584	1.4796	6.5083
1.5117	1.2859	0.8188	3.2250	1.3022	6.6134
4.7780	1.8619	0.1939	0.5209	1.7917	6.6288

Applying the proposed model, the optimal solution was $\tau_1 = 3$ and $\tau_2 = 53$ with $\hat{\beta} = 1.1$ and $\hat{\theta} = 1.8$. The model predicted the end of the useful life to be at 6.4703 years which is close to the actual end of the useful life of 6.825 years. It is worth to mention here that the model incorrectly predicted a small infant mortality region.This error is reasonable as the predicted infant mortality region is very small and contains only two data points. Moreover, the predicted start of the useful life region was 0.1143 years which is reasonable close to the actual value.

D.Example 4

Table IV shows the pertaining failure time data for example 4. The data for this example is the same data used in examples 1, 2, and 3 together.

TABLE IV
FAILURE TIME DATA FOR EXAMPLE 4

0.0039	2.0695	3.285	0.0966	0.006	2.0807	6.4353
0.0999	0.6825	0.1272	2.0023	1.1071	1.5107	6.3657
0.5016	6.825	0.3807	0.8319	1.8794	0.869	6.607
0.1743	2.7308	0.7918	0.9206	0.7167	0.3098	6.4703
0.175	5.5971	0.1143	2.6753	1.4214	1.6911	6.6311
0.1047	0.3631	0.3518	0.1397	0.4155	2.7609	6.5353
0.1009	0.1321	1.2889	1.3137	1.2121	0.1705	6.3927
0.6425	3.1936	5.2825	0.9347	0.9584	1.4796	6.5083
1.407	1.5117	1.2859	0.8188	3.225	1.3022	6.6134
0.2395	4.778	1.8619	0.1939	0.5209	1.7917	6.6288

Applying the proposed model, the optimal solution was $\tau_1 = 10$ and $\tau_2 = 65$ with $\hat{\beta} = 1.06$ and $\hat{\theta} = 2$. Fig. 2 shows the three regions found by the model in this example.

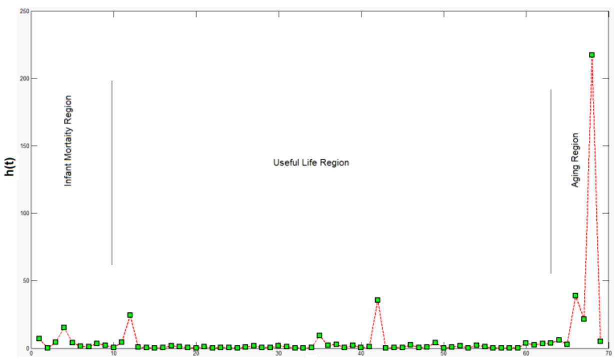

Fig. 2 The bathtub regions found by the model for example 4

The model was able to distinguish between three regions. The infant mortality region that contains 10 data points, the useful life region between 0.1397 years and 6.5353 years that contains 55 data points, and the aging region that contains 5 data points. It is obvious that the model incorrectly assign the beginning and the end of the useful life region but with a reasonable error as the actual beginning and ending for the useful life region are 0.006 and 6.825 year respectively.

IV. CONCLUSIONS

In this paper we presented a practical model utilizing non-parametric analysis of ungrouped complete failure time data for Weibull-life PV systems. The illustrative examples showed that the proposed model reasonably distinguished between the infant mortality, useful life, and aging regions of Weibull-life PV systems. Moreover, the results for the illustrative examples showed that the model provided a reasonable estimation of the life distribution parameters in the useful life region.

A good estimation for the life distribution parameters of the PV systems will allow the decision maker to incorporate the life distribution of the PV systems in their calculations when making decisions regarding the feasibility of a certain PV system as this information will enhance the calculations' accuracy of the payback period and the energy price per KWh.

ACKNOWLEDGMENT

The author is grateful to the Applied Science Private University, Amman, Jordan, for the financial support granted to this research (Grant No.DRGS-2015-2014).

REFERENCES

[1] E. F. Hitt, Battelle and O. H. Columbus, "Total ownership cost use in management", Digital Avionics Systems Conference, Proceedings, 17th DASC. The AIAA/IEEE/SAE, Vol. 1, A32-1-5,(1998).
[2] Maish A, Atcitty C., Hester D. Greenberg D., Osborn D., Collier D., (1997). Photovoltaic Reliability, Proceedings of the 26th Photovoltic Specialists Conference (PVSC) Anheim CA 1049.
[3] Begovic M., Pregelj A., Rohatgi A., (2000). Four-year Performance Assessment of the 342 kW PV System at Georgia Tech, Proceedings of the 28th Photovoltic Specialists Conference (PVSC) Anchorage AL.
[4] Guess, F. M., Usher, J. S., Hodgson, T. J. (1991). Estimating system and component reliabilities under partial information on cause of failure. J. Statist. Plann. Infer.29:75–85.
[5] Flehinger, B. J., Reiser, B., and Yashchin, E. (1998). Survival with competing risks and masked causes of failures. Biometrika 85:151–164.
[6] Basu, S., Basu, A. P., and Mukhopadhyay, C. (1999). Bayesian analysis for masked system failure data using non-identical weibull models. J. Statist. Plann. Infer. 78:255–275.
[7] Chiranjit M. and Sanjib B., (2007). Bayesian Analysis of Masked Series System Lifetime Data. Communications in Statistics—Theory and Methods, 36: 329–348.
[8] Charles E., (2008). An Introduction to Reliability and Maintainability Engineering,McGraw-Hill, New Delhi.
[9] Elsayed A., (1996). Reliability Engineering, Addison Wesley, Massachusetts.

Saleem Z. Ramadan is an Assistant Professor in the Department of Mechanical and Industrial Engineering at Applied Science University, ShafaBadran, Jordan. He received his Ph.D. from the Department of Industrial and Systems Engineering at Ohio University in 2011. His research interests are in Bayesian statistics, reliability, genetic algorithms, and operations research.