Estimating Component Reliabilities from Incomplete System Failure Data

Huairui Guo, Ph.D, ReliaSoft Corporation

Ferenc Szidarovszky, Ph.D, ReliaSoft Corporation

Pengying Niu, ReliaSoft Corporation

Key Words: masked failure data, maximum likelihood estimate, series systems, parallel systems

SUMMARY & CONCLUSIONS

Engineers often need to estimate component reliabilities from system failure data that are obtained either from in-house tests or field operations. The components that caused the system failures usually can be identified by checking the failed systems; however, in some cases, the exact cause of failure is very difficult or impossible to identify. For the latter case, the information on the system failures is not complete, and such data are usually called masked failure data. In this paper, we propose a method for estimating component reliabilities from masked system failure data. Solutions for systems with serial and parallel configurations are provided. The failure distribution of each component is obtained from the proposed method, and is then used to calculate the probability that a system failure is caused by the given component when the exact time or an interval time of the system failure is known. This probability is very useful since it can be used to decide which component should be analyzed first when a system failure occurs, depending on the failure time.

1 INTRODUCTION

Estimating the system reliability from component test data has been discussed by many researchers and engineers. If the system reliability diagram and the component test data are available, then the system reliability can be estimated easily since it is a function of the component reliabilities. However, in many other applications, the component reliabilities need to be estimated from system failure data. Sometimes, it is not easy or even possible to identify the components that caused the system failure. For example, consider a series system with 3 components, where the failure of the system can be caused by components 1, 2, or 3. When the exact cause of system failure can be identified, for instance, the failure was caused by component 1, then the system failure time is the failure time of component 1 and also the suspension time for components 2 and 3 since these two components were still working when the system failed (in this paper, we assume that components cannot fail at the exact same time). When the cause can only be identified by a subset of components, for example the failure is caused by either 1 or 2, but not by 3, then the system failure time is the possible failure time for 1 or 2, and it is also the suspension time for component 3. The following table illustrates these two scenarios.

Table 1. Scenarios for Masked Failure Data

System ID	Failure/Suspension Time	Cause of Failure
1	10	{1}
2	4	{2}
3	12	{1, 2}
4	45	{2, 3}
5	5	{3}
6	78	{1, 2, 3}
7	100	Suspended

There are 7 systems in Table 1. The notation {1, 2} means that the system failure is caused by either component 1 or component 2. The last system is suspended at time 100, which means that 100 is the suspension time for all the components.

Table 1 shows only data with exact failure times and suspensions. Left censored and interval failure data may also be observed. Masked failure data also occurs for parallel systems. Therefore, estimating the component reliability from masked system failure data is a challenging task.

Different methods have been proposed to estimate component reliabilities for series systems. Miyakawa [1] gave a MLE (maximum likelihood estimation) solution for series systems having two components with exponential distributions (constant failure rate). User and Hodgon [2] applied Miyakawa's method to series systems of three components, again assuming that each component follows an exponential distribution. They later [3] extended their method to series systems with n components and provided the exact MLE solution. The methods in [1, 2, 3] work only for exponential distribution components. To overcome this, User [4] presented a MLE method for series systems consisting of components with Weibull distributions. In recent years, Bayes analysis [5, 6] was integrated with the MLE method proposed in [1, 2, 3]. Although intensive study has been conducted for series systems, little research has been done for parallel systems. Tan [7] proposed an equivalent failure time approach for series and parallel systems with exponential components. Sarhan and El-Bassiouny [8] used MLE and Bayes analysis for parallel systems with exact failure times. However, no research has

978-1-4673-4711-2/13/$31.00 ©2013 IEEE

been found for masked data with interval failure times and suspensions.

In this paper, we will provide a method for estimating component reliability from masked system failure data. This method is based on the MLE theory, and it is a general method. It can be used for:

- Systems with components of any distributions, not only of exponential distribution.
- Both series and parallel systems.
- Different data types including exact failure time, right censored, left censored, and interval failure data.

The paper is organized as follows: Section 2 introduces the background on MLE. Section 3 presents the modeling strategy and an example for series systems. Some statistical proprieties of the cause of failures are also discussed. Parallel systems are analyzed in section 4.

2 BACKGROUND ON MLE

MLE is probably the most popular parameter estimation method used in life data analysis. Since the proposed method in this paper is based on MLE theory, we will briefly explain how to use MLE in life data analysis.

In general, there are three types of life data: exact failure, interval failure, and suspensions (also called right censored data). They are illustrated in Figure 1.

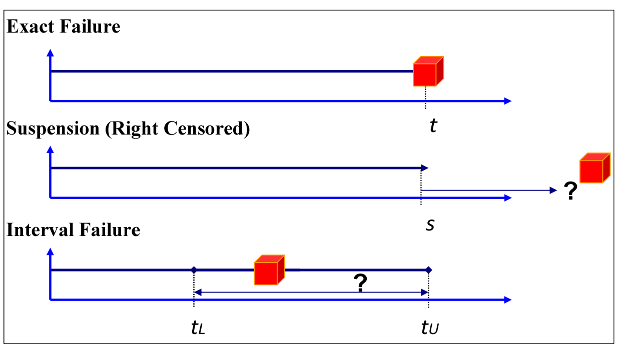

Figure 1. Three Types of Life Data

From Figure 1 we can see that for a failure of a product, if its failure time t is known, it is called an exact failure. If the exact failure time is unknown, but we know that the failure occurred within a time interval between t_1 and t_2, then this failure is called interval failure. If the system is still running at time s when the data are collected, then s is called the suspension time. Different likelihood functions are built for these three different data types. The likelihood function for the ith observation t_i is:

$$l_i = \begin{cases} f(t_i) & \text{if the } i\text{th observation is an exact failure} \\ F(t_{i,U}) - F(t_{i,L}) & \text{if the } i\text{th observation is an interval failure} \\ R(s_i) & \text{if the } i\text{th observation is a suspension} \end{cases} \quad (1)$$

where $f(t_i)$ is the probability density function (*pdf*) for the failure time, $F(t_i)$ is the probability of failure or cumulative distribution function (*cdf*), and $R(s_i)$ is the reliability function. When there are n observations, the complete likelihood function is:

$$L = \prod_{i=1}^{n} l_i \quad (2)$$

and the logarithm transform of the likelihood function is:

$$\ln(L) = \Lambda = \sum_{i=1}^{n} \ln(l_i) \quad (3)$$

In order to use MLE, a distribution has to be assumed for the failure time, and then parameters in the distribution can be estimated by maximizing the log-likelihood value in equation (3). For example, assuming a Weibull distribution, the log-likelihood function for a data set with all the three data types would be:

$$\Lambda = \sum_{i=1}^{n_e} \ln\left[\frac{\beta}{\eta}\left(\frac{t_i}{\eta}\right)^{\beta-1} e^{-\left(\frac{t_i}{\eta}\right)^{\beta}}\right] + \sum_{i=1}^{n_I} \ln\left[e^{-\left(\frac{t_{i,U}}{\eta}\right)^{\beta}} - e^{-\left(\frac{t_{i,L}}{\eta}\right)^{\beta}}\right] - \sum_{i=1}^{n_s}\left(\frac{s_i}{\eta}\right)^{\beta} \quad (4)$$

where β, η are the distribution parameters; n_e, n_I, and n_s are the numbers for exact failures, interval failures, and suspensions, respectively in the observation. The parameters β, η can be obtained by maximizing equation (4). Equation (4) provides only point estimates for β, η. To get the confidence interval of β, η, we can either calculate the variance and covariance matrix of β, η through the Fisher information matrix or use the likelihood ratio statistic. For more details on MLE and confidence intervals of distribution parameters, the readers can be referred to [9, 10].

3 METHODOLOGY AND EXAMPLE FOR SERIES SYSTEMS

The MLE method given in section 2 can be applied to both component and system failures. For a series system, system failure occurs whenever a component fails. Each component represents a competing failure mode. The system reliability function is:

$$R_s(t) = \prod_{j=1}^{k} R_j(t) \quad (5)$$

and the probability of failure is:

$$F_s(t) = 1 - R_s(t) \quad (6)$$

Taking the derivative of $F_s(t)$ with respect to time t, the system *pdf* can be obtained as:

$$f_s(t) = \sum_{j=1}^{k}\left(f_j(t)\prod_{p\neq j}^{k} R_p(t)\right) \quad (7)$$

Using equations (5) to (7), we can estimate the system failure distribution based on the MLE knowledge in section 2.

To estimate the component failure distributions from the system failures, we divide the system failures into two categories: failures with a clear cause and failures with masked causes. The likelihood functions are established next for these two categories.

3.1 Failures with A Clear Cause

For this type of failure, the component that caused a system failure can be identified. If the *i*th failure is an exact failure and the *j*th component is the cause, then the likelihood function is:

$$l_i = f_j(t_i)\prod_{p\neq j}^{k} R_p(t_i) \tag{8}$$

where t_i is the failure time, $f_j(t_i)$ is the *pdf* of the *j*th component, $R_p(t_i)$ is the reliability of the *p*th component, and *k* is the total number of components in the series system.

If the *i*th observation is an interval failure and the *j*th component is identified as the exact cause, then the likelihood function is:

$$l_i = \int_{t_{i,L}}^{t_{i,U}} f_j(t)\prod_{p\neq j} R_p(t)dt \tag{9}$$

where $t_{i,U}$ and $t_{i,L}$ are the upper and the lower bounds of the interval.

3.2 Failures with Masked Causes

If the component that caused the system failure cannot be clearly identified, then we will end up with masked failures that are similar to the example in Table 1. Define C_i as the set of components that is possibly causing the *i*th failure, and $\overline{C}_i$ as the set containing all other components. If only one component is in C_i, then it becomes the case where the cause can be clearly identified, as described in section 3.1. If C_i has all the components, then the cause of the failure is completely masked. The likelihood function for exact failures and interval failures with masked causes are given below.

If the *i*th failure is an exact failure and t_i is the failure time, then the likelihood function is:

$$\begin{aligned} l_i &= \sum_{j\in C_i}\left[f_j(t_i)\prod_{\substack{p\in C_i\\ p\neq j}} R_p(t_i)\right]\times\prod_{p\in \overline{C}_i} R_p(t_i) \\ &= \sum_{j\in C_i}\left[f_j(t_i)\prod_{p\neq j} R_p(t_i)\right] \end{aligned} \tag{10}$$

If the *i*th failure is an interval failure, then the likelihood function is:

$$l_i = \sum_{j\in C_i}\left[\int_{t_{i,L}}^{t_{i,U}} f_j(t)\prod_{p\neq j} R_p(t)dt\right] \tag{11}$$

3.3 Likelihood Function for Suspensions

If the *i*th observation is a suspension, then it is clear that the likelihood function is

$$l_i = R_s(t_i) = \prod_{p=1}^{k} R_p(t_i) \tag{12}$$

The complete log-likelihood function can be calculated using equation (3). The MLE solution of the parameters for the failure time distribution of each component can be obtained by maximizing the complete likelihood function. Since the complete likelihood function is a continuous function, any gradient-based optimization algorithm can be used. If some of the components in the system are the same, then a common distribution is needed for those components.

3.4 An Example for Series Systems

Assume that thirty identical systems are tested to failure and that the failure information is collected in Table 2. Each system has three components in series.

Table 2. Example Data for a Series System

System ID	Failure Time	Causes	System ID	Failure Time	Causes
1	21	{2}	16	281	{1}
2	38	{1,2}	17	295	{2}
3	54	{3}	18	310	{3}
4	66	{3}	19	338	{3}
5	76	{1,2}	20	341	{2}
6	78	{2,3}	21	354	{1}
7	123	{3}	22	358	{2}
8	130	{1,3}	23	431	{1,2,3}
9	152	{1,2,3	24	457	{3}
10	159	{1}	25	545	{1,2,3}
11	199	{3}	26	569	{2}
12	201	{1}	27	677	{3}
13	204	{1}	28	818	{2}
14	215	{2,3}	29	946	{2,3}
15	218	{1,2}	30	1486	{1}

Solution:

Assume that each component follows a Weibull distribution. We can use equations (8) to (11) to get the likelihood function for each observation, depending on the failure type and information on the causes. The results are $\beta_1 = 1.2576,\ \eta_1 = 994.3661$, $\beta_2 = 1.1635,\ \eta_2 = 908.9458$, and $\beta_3 = 1.1308,\ \eta_3 = 840.1141$. The final likelihood value is -228.6851. Using the estimated component distributions, we can predict the system reliability. The observed and predicted probabilities of the system failure (unreliability), as functions of *t*, are given in Figure 2.

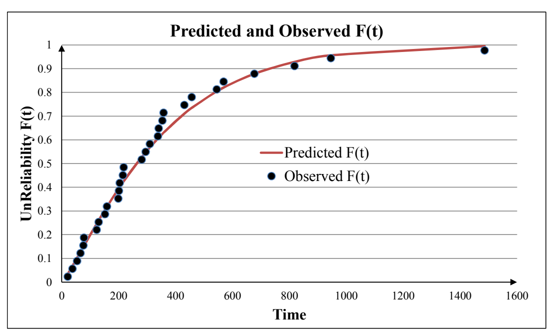

Figure 2. Plot of Unreliability F(t) vs. Time for a Series System

The observed *F*(*t*) values in Figure 2 are calculated using median rank.

3.5 Probability of A System Failure Caused by A Given Component

Once the distribution of each component is obtained, we can calculate the probability that the failure was caused by a particular component, given that the system failed at time t. The probability of failure caused by the jth component is:

$$P_j = \frac{f_j(t)\prod_{p\neq j} R_p(t)}{\sum_{i=1}^{k}\left(f_i(t)\prod_{p\neq i} R_p(t)\right)} = \frac{f_j(t)/R_j(t)}{\sum_{i=1}^{k}\left(f_i(t)/R_i(t)\right)} = \frac{h_j(t)}{\sum_{i=1}^{k} h_i(t)} \quad (13)$$

where $h_j(t)$ is the failure rate (hazard rate) for the jth component at time t. It is clear that P_j is the ratio of the likelihood value of the failure caused by component j and the sum of the likelihood values of the failure caused by all the components.

Similarly, for a system failure that occurred within time t_L and t_U, the probability of failure caused by the jth component is also the likelihood ratio as:

$$P_j = \frac{\int_{t_L}^{t_U} f_j(t)\prod_{p\neq j} R_p(t)dt}{\sum_{i=1}^{k}\left(\int_{t_L}^{t_U} f_i(t)\prod_{p\neq i} R_p(t)dt\right)} \quad (14)$$

Clearly, equations (13) and (14) also can be used for failures with masked causes when the subset (subsystem) with the possible failed component is treated as a "system." With the calculated probabilities of equations (13) and (14), failure analysis can be conducted first on the components that have a high probability of causing the system failure. This saves time and money.

4 METHODOLOGY AND EXAMPLE FOR PARALLEL SYSTEMS

Similar to series systems, we can also get the likelihood function for a parallel system and use MLE to get the failure distribution of each component. For a parallel system, system failure occurs when all the components have failed. However, the system will work as long as there is at least one working component; therefore, the reliability function for a parallel system with k components is:

$$R_s(t) = 1 - \prod_{j=1}^{k} F_j(t) = 1 - \prod_{j=1}^{k}\left(1 - R_j(t)\right) \quad (15)$$

where $F_j(t)$ is the *cdf* of failure time for the jth component.

The probability of failure of the system is:

$$F_s(t) = \prod_{j=1}^{k} F_j(t) \quad (16)$$

Taking the derivative of $F_s(t)$ with respect to t, the *pdf* of the system failure time is:

$$f_s(t) = \sum_{j=1}^{k}\left(f_j(t)\prod_{p\neq j} F_p(t)\right) \quad (17)$$

Using equations (15) to (17), we can estimate the system failure distribution. The following sections discuss how to estimate the component reliabilities for a parallel system.

4.1 Failures with A Clear Cause

For a parallel system, we define the last failed component as the cause for a system failure. If the last failed component can be identified, then it is a failure with a clear cause. If the ith observation is an exact failure and the jth component is the cause, then the likelihood function for this observation is:

$$l_i = f_j(t_i)\prod_{p\neq j} F_p(t_i) \quad (18)$$

where $F_p(t_i)$ is the probability of failure for the pth component at the ith observation time t_i.

If the ith observation is an interval failure and the jth component is identified as the cause, then the likelihood function is:

$$l_i = \int_{t_{i,L}}^{t_{i,U}} f_j(t)\prod_{p\neq j} F_p(t)dt \quad (19)$$

where $t_{i,U}$ and $t_{i,L}$ are the upper and the lower values of the interval.

4.2 Failures with Masked Causes

Using the same notation for the set of masked causes, as in section 3.2, the likelihood function for exact failures and interval failures with masked causes are given below.

If the ith failure is an exact failure and t_i is the failure time, the likelihood function is:

$$\begin{aligned} l_i &= \sum_{j\in C_i}\left[f_j(t_i)\prod_{\substack{p\in C_i \\ p\neq j}} F_p(t_i)\right]\times \prod_{p\in \bar{C}_i} F_p(t_i) \\ &= \sum_{j\in C_i}\left[f_j(t_i)\prod_{p\neq j} F_p(t_i)\right] \end{aligned} \quad (20)$$

If the ith failure is an interval failure, then the likelihood function is:

$$l_i = \sum_{j\in C_i}\left[\int_{t_{i,L}}^{t_{i,U}} f_j(t)\prod_{p\neq j} F_p(t)dt\right] \quad (21)$$

4.3 Likelihood Function for Suspensions

If the ith observation is a suspension, then clearly the likelihood function is

$$l_i = R_s(t_i) = 1 - \prod_{p=1}^{k}\left(1 - R_p(t_i)\right) \quad (22)$$

The complete log-likelihood function can be calculated using equation (3). The MLE solution of the parameters for the failure time distribution of each component can be obtained by maximizing the complete likelihood function.

4.4 An Example for Parallel Systems

Assume that fifteen systems are tested to failure and that the failure information is collected in Table 3. Each system has three components in parallel.

Table 3. Example Data for a Parallel System

System ID	Failure Time	Causes	System ID	Failure Time	Causes
1	161	{1}	9	246	{1, 3}
2	232	{2}	10	38	{2}
3	379	{3}	11	61	{3}
4	159	{1, 2}	12	259	{1}
5	94	{2, 3}	13	231	{1, 2}
6	133	{2}	14	22	{2, 3}
7	91	{1}	15	39	{1, 2, 3}
8	231	{1, 2, 3}			

Solution:

Assume that each component follows a Weibull distribution. We can use equations (18) and (20) to get the likelihood function for each observation. The results are $\beta_1 = 0.8806$, $\eta_1 = 87.7259$, $\beta_2 = 1.1168$, $\eta_2 = 88.7596$, and $\beta_3 = 0.6263$, $\eta_3 = 45.6395$. The final likelihood value is -99.6779. Using the estimated component distribution, we can predict the system reliability. The observed and predicted probabilities of the system failure (unreliability), as functions of t, are given in Figure 3.

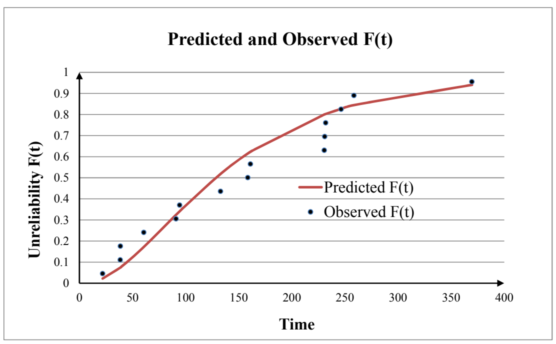

Figure 3. Plot of Unreliability F(t) vs. Time for a Parallel System

The observed $F(t)$ values in Figure 3 are calculated using median rank.

4.5 Probability of A System Failure Caused by A Given Component

The following equations can be used to calculate the probability that a component is the last failed component, given that the system (or a subset of the system) has failed. For an exact failure at time t, the probability that the jth component is the last failed component is:

$$P_j = \frac{f_j(t)\prod_{p\neq j} F_p(t)}{\sum_{i=1}^{k}\left(f_i(t)\prod_{p\neq i} F_p(t)\right)} = \frac{f_j(t)/F_j(t)}{\sum_{i=1}^{k}\left(f_i(t)/F_i(t)\right)} \quad (22)$$

For an interval failure occurring within time t_L and t_U, the probability that the jth component is the last failed component is:

$$P_j = \frac{\int_{t_L}^{t_U} f_j(t)\prod_{p\neq j} F_p(t)dt}{\sum_{i=1}^{k}\left(\int_{t_L}^{t_U} f_i(t)\prod_{p\neq i} F_p(t)dt\right)} \quad (23)$$

These calculated probabilities can help engineers make decisions on failure analysis.

5 CONCLUSIONS

In this paper, a general MLE method on estimating component reliabilities from masked system failures is proposed. Most of the existing methods apply only to the exponential distribution with exact failure data. The proposed method can be used for all data types and any distributions. Two examples are provided to show the ease of use of the proposed method. In both examples, the predicted system unreliabilities match the observed values very well. This shows that the proposed method is valid. The MLE method in this paper also can be integrated with Bayes theory to estimate component reliabilities, if there is prior information on the components [5, 6].

REFERENCES

1. M. Miyakawa, "Analysis of incomplete data in competing risks model," *IEEE Transaction on Reliability*, vol. R-33, no. 4, 1984, pp. 293-296.
2. J. User and T. Hodgson, "Maximum likelihood analysis of component reliability using masked system life-test data," *IEEE Transaction on Reliability*, vol. 37, no. 5, 1988, pp. 550-555.
3. D. Lin, J. User, F. Guess, "Exact maximum likelihood estimation using masked system data," *IEEE Transaction on Reliability*, vol. 42, no. 4, 1993, pp. 631-635.
4. J. User, "Weibull component reliability prediction in the presence of masked data," *IEEE Transaction on Reliability*, vol. 45, no. 2, 1996, pp. 229-232.
5. D. Lin, J. User, F. Guess, "Bayes estimation of component reliability from masked system life data," *IEEE Transaction on Reliability*, vol. 45, no. 2, 1996, pp. 233-237.
6. A. Sarhan, "Reliability estimations of components from masked system life data," *Reliability Engineering & System Safety*, vol. 74, 2001, pp.107-113.
7. Zhibin Tan, "Estimation of exponential component reliability from uncertain life data in series and parallel systems," *Reliability Engineering & System Safety*, vol. 92, 2007, pp.223-230.
8. A. Sarhan, A. El-Bassiouny, "Estimation of components reliability in a parallel system using masked system life data," *Applied Mathematics and Computation*, vol. 138, no. 1, pp. 61-75.
9. N. Doganaksoy, "Interval estimation from censored & masked system failure data," *IEEE Transaction on Reliability*, vol. 40, no. 3, 1991, pp. 280-286.
10. W. Q. Meeker and L. A. Escobar, *Statistical Methods for Reliability Data*, John Wiley & Sons, Inc., New York, 1998.

BIOGRAPHIES

Huairui Guo
ReliaSoft Corporation
1450 S. Eastside Loop
Tucson, AZ, 85710

e-mail: Harry.Guo@ReliaSoft.com

Dr. Huairui (Harry) Guo is the Director of the Theoretical Development Department at ReliaSoft Corporation. He received both his Ph.D. in Systems & Industrial Engineering and M.S. in Reliability & Quality Engineering from the University of Arizona. His research and publications cover reliability areas, such as life data analysis, repairable system modeling and reliability test planning, and quality areas, such as process monitoring, analysis of variance, and design of experiments. In addition to research and product development, he is also part of the training and consulting arm and has been involved in various projects from the automobile, medical device, oil and gas, and aerospace industries. He is a certified reliability professional (C.R.P), ASQ certified CQE. He is a member of IIE, SRE, and ASQ.

Ferenc Szidarovszky
ReliaSoft Corporation
1450 S. Eastside Loop
Tucson, AZ, 85710

e-mail: Ferenc.Szidarovszky@Reliasoft.com

Dr. Ferenc Szidarovszky received his education in Hungary, where he earned two PhD degrees, one in mathematics, the other in economics. The Hungarian Academy of Sciences has awarded him with the degrees of "Candidate in Mathematics" and "Doctor of Engineering Sciences." Between 1973 and 1986 he was an investigator in several joint research projects between Hungarian and American researchers, and in 1988 he became a professor in the Systems and Industrial Engineering Department of the University of Arizona. In 2011 he retired from the university and joined ReliaSoft as a senior researcher. His research interests are decision making under uncertainty, multi-criteria decision making, game theory, conflict resolution and their applications in natural resource management, industry, economics and cyber security. He is the author of 18 books and over 300 refereed journal publications in addition to numerous conference presentations and invited lectures. He is regularly invited to give short courses in Game Theory in several countries in Europe and Asia.

Pengying Niu
ReliaSoft Corporation
1450 S. Eastside Loop
Tucson, AZ, 85710

e-mail: Pengying.Niu@ReliaSoft.com

Pengying Niu is a research scientist at ReliaSoft Corporation. She is currently playing a key role in the development of Lambda Predict. Before joining ReliaSoft, she worked at Texas Instruments where she was involved in IC design and testing. She received her master’s degree from the National University of Singapore, and M. E. in Electrical and Computer Engineering from the University of Arizona. She has done extensive work on AC/DC and DC/AC converters. Her current research interests include reliability prediction and physics of failure for electronic components such as MOSFET, IGBT and electronic systems. She is also an ASQ Certified Reliability Engineer (CRE) and a Certified Quality Engineer (CQE).