

A Tukey's g-and-h distribution based approach with PSO for degradation reliability modeling

Degradation
reliability
modeling

1699

Jinbao Zhang
Harbin Institute of Technology, Harbin, China

Yongqiang Zhao
*School of Mechatronic Engineering, Harbin Institute of Technology,
Harbin, China, and*

Ming Liu and Lingxian Kong
Harbin Institute of Technology, Harbin, China

Received 6 November 2017
Revised 13 February 2019
Accepted 13 February 2019

Abstract

Purpose – A generalized distribution with wide range of skewness and elongation will be suitable for the data mining and compatible for the misspecification of the distribution. Hence, the purpose of this paper is to present a distribution-based approach for estimating degradation reliability considering these conditions.

Design/methodology/approach – Tukey's g-and-h distribution with the quantile expression is introduced to fit the degradation paths of the population over time. The Newton-Raphson algorithm is used to approximately evaluate the reliability. Simulation verification for parameter estimation with particle swarm optimization (PSO) is carried out. The effectiveness and validity of the proposed approach for degradation reliability is verified by the two-stage verification and the comparison with others' work.

Findings – Simulation studies have proved the effectiveness of PSO in the parameter estimation. Two degradation datasets of GaAs laser devices and crack growth are performed by the proposed approach. The results show that it can well match the initial failure time and be more compatible than the normal distribution and the Weibull distribution.

Originality/value – Tukey's g-and-h distribution is first proposed to investigate the influence of the tail and the skewness on the degradation reliability. In addition, the parameters of the Tukey's g-and-h distribution is estimated by PSO with root-mean-square error as the object function.

Keywords PSO, Quantile regression, Degradation path, Reliability estimation, Tukey's g-and-h distribution

Paper type Research paper

1. Introduction

Reliability is a significant indicator to the performance estimation and the quality control of products. For highly reliably products, failure data are uneasy to get even in the accelerated

© Jinbao Zhang, Yongqiang Zhao, Ming Liu and Lingxian Kong. Published by Emerald Publishing Limited. This article is published under the Creative Commons Attribution (CC BY 4.0) licence. Anyone may reproduce, distribute, translate and create derivative works of this article (for both commercial and non-commercial purposes), subject to full attribution to the original publication and authors. The full terms of this licence may be seen at <http://creativecommons.org/licenses/by/4.0/legalcode>

The authors would like to thank the editors and the anonymous reviewers for their useful comments which have greatly improved the paper. This work is supported by the National Natural Science Foundation of China (No. 51505100).



Engineering Computations
Vol. 36 No. 5, 2019
pp. 1699-1715
Emerald Publishing Limited
0264-4401
DOI 10.1108/EC-11-2017-0428

test. In this case, degradation data will be an efficient alternative for reliability estimation. The existing models for degradation reliability in detail have been summarized (Gorjian *et al.*, 2009; Ye and Xie, 2015). A common approach is to extrapolate the degradation paths to the failure threshold and obtain the pseudo-lifetime data for reliability estimation (Lu and Meeker, 1993). The other way to estimate reliability is based on the relationship between the degradation data and the failure threshold (Zuo *et al.*, 1999). The relationship between models based on the degradation data and models based on the pseudo-lifetime data are stated by Meeker and Escobar (1998). However, the present models for degradation reliability are based on normal distribution, Weibull distribution and so on, and misspecification of distributions may exist. Hence, a generalized distribution is required.

The g-and-h distribution generated by the transformation of the standard normal distribution, is first proposed by Tukey (1977). Basic properties are studied (Jorge and Boris, 1984; Hoaglin *et al.*, 1985; Macgillivray, 1992) and the modifications are extensively investigated (Morgenthaler and Tukey, 2000; Fischer, 2010). It has extensive range of skewness and elongation and corresponding control parameters, which could depict the properties of data in detail. Comparison of goodness-of-fit has been performed among gh-transformed distributions and other distributions, and the results proved that in most cases the original Tukey's g-and-h distribution would provide a better fit (Klein and Mittnik, 2002). In practice, Tukey's g-and-h distribution has been widely used in various fields, such as common stock return (Badrinath and Chatterjee, 1991), modeling extreme wind speeds (Dupuis and Field, 2004), operational risk (Degen *et al.*, 2007), value-at-risk (Jiménez and Arunachalam, 2011), and simulation studies (He and Raghunathan, 2006). Nevertheless, it has been not investigated in reliability yet.

The advantages of Tukey's g-and-h distribution for fitting degradation paths are summarized as follows:

- It has wide range of skewness and kurtosis which could be robust with extreme values and outliers, and can judge whether the distribution is heavy tail.
- It is expressed by a quantile regression and can be easily introduced into the location-scale model, which will well deal with heteroscedasticity over time.
- It can accommodate a rich class of common univariate distributions, such as normal, lognormal, Weibull and exponential (Morgenthaler and Tukey, 2000; Chaudhuri and Ghosh, 2016).

One prominent problem lies in the parameter estimation of Tukey's g-and-h distribution, especially for the degradation situation considering time. There are mainly three methods for parameter estimation, which are quantile method (Hoaglin *et al.*, 1985), method of moments (Headrick *et al.*, 2008), and maximum likelihood method (Majumder and Ali, 2008), respectively. Maximum likelihood estimation (MLE) is the most popular approach; nevertheless, its likelihood function cannot be written in closed form and finally induce the intractable computation. Therefore, numerical methods are introduced to solve the MLE by approximation, which may be complicated (Rayner and Macgillivray, 2002a; Xu and Genton, 2015; Bee and Trapin, 2016). Method of moments is another approach for parameter estimation, but it may be not appropriate for the situation varying with time.

Quantile-based estimation method would be a natural alternative for Tukey's g-and-h distribution. A weighted quantile-based estimation shows better effects in a number of ways than the numerical MLE for smaller and moderately-sized samples (Rayner and Macgillivray, 2002b). In addition, a rapidly convergent quantile based least squares estimation method (Xu *et al.*, 2014) has been developed to fit the g-and-h distributional family parameters. However, when modeling the degradation path, more parameters

including g and h should be estimated. In order to overcome this difficulty, particle swarm optimization (PSO) (Kennedy and Eberhart, 1995) as a heuristic optimization method is brought in for quantile-based estimation. PSO has many advantages including simplicity of implementation, robust, and has been successfully used in data-driven fuzzy clustering (Chen and Zhao, 2009), solving the likelihood objective function of 3-p Weibull distribution for parameter estimation (Örkcü *et al.*, 2015).

In this paper, emphasize is put on the reliability estimation by Tukey’s g -and- h distribution for the degradation data. The rest of the paper is organized as follows. In Section 2, Tukey’s g -and- h distribution and its properties are presented, the influence of parameters on the distributions are investigated as well. In addition, the quantile-based estimation is introduced in detail. Section 3 employs an intelligent optimization method called PSO for the parameter estimation. The practical applications for degradation reliability estimation concerning GaAs laser device and crack growth are performed in Section 4. Conclusion is given in Section 5.

2. Tukey’s g -and- h distribution

Definition Let Z be a random variable with the standard normal distribution, then a random variable U defined by:

$$U = q_{g,h}(Z) = g^{-1}[\exp(gZ) - 1]\exp(hZ^2/2) \tag{1}$$

is said to have Tukey’s g -and- h distribution. $q_{g,h}(z)$ should be a strictly increasing monotonic function of z , that is, $q'_{g,h}(z) > 0$ with parameters $g \in R$ and $h \geq 0$. When $h < 0$, the transformation is no longer strictly monotone increasing and induces a finite support of data (Klein and Mitnik, 2002), so in this paper, the parameter h will be constrained in the interval $[0, \infty)$.

Two subclasses of distributions based on equation (1) are the g and the h classes which are defined as follows:

$$U = q_{g,0}(Z) = \lim_{h \rightarrow 0} q_{g,h}(Z) = g^{-1}\{\exp(gZ) - 1\} \tag{2}$$

$$U = q_{0,h}(Z) = \lim_{g \rightarrow 0} q_{g,h}(Z) = z \exp(hZ^2/2) \tag{3}$$

When $g = h = 0$, $U = Z$, so U has a standard normal distribution. Furthermore, the standard normal distribution could be replaced by other symmetric distributions, such as Laplace distribution, generalized error distribution and so on (Morgenthaler and Tukey, 2000; Klein and Mitnik, 2002).

2.1 Properties

Some properties are introduced for the following verification. The probabilistic density function (pdf) of standard normal distribution is:

$$\varphi_Z(z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) \tag{4}$$

and its cumulative density function (cdf) is:

$$\Phi_Z(z) = \Pr(Z < z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z \exp\left(-\frac{w^2}{2}\right) dw, \quad -\infty < z < +\infty. \tag{5}$$

The formulae for k -th moment of the g -and- h distribution is:

$$E[q(z)^k] = \int_{-\infty}^{+\infty} q(z)^k f_Z(z) dz \tag{6}$$

Then the mean of Tukey's g -and- h distribution is:

$$\mu_{g,h} = E[q_{g,h}(z)] = \begin{cases} \frac{1}{g\sqrt{1-h}} \left[\exp\left(\frac{1}{2} \frac{g^2}{1-h}\right) - 1 \right], & g \neq 0, 0 \leq h < 1 \\ 0, & g = 0 \end{cases} \tag{7}$$

and the variance:

$$\sigma_{g,h}^2 = D[q_{g,h}(z)] = E[q_{g,h}(z)^2] - E^2[q_{g,h}(z)] \tag{8}$$

where

$$E[q_{g,h}(z)^2] = \frac{1 - 2\exp\{g^2/(2 - 4h)\} + \exp\{2g^2/(1 - 2h)\}}{g^2(1 - 2h)^{1/2}} \tag{9}$$

2.2 Relationship with reliability

Since the function

$$u = q_{g,h}(z) = g^{-1}(e^{gz} - 1)e^{hz^2/2} \tag{10}$$

for $h > 0$ is strictly increasing, the cdf of a g -and- h random variable U can be written as:

$$F(u) = F(z) = \Phi[q_{g,h}^{-1}(u)] \tag{11}$$

where $z = q_{g,h}^{-1}(u)$, which can be solved with the Newton-Raphson method, that is:

$$z^{(n_0+1)} = z^{(n_0)} - f(z^{(n_0)})/f'(z^{(n_0)}) \tag{12}$$

in which n_0 is the iterative number, and by transformation of [equation \(10\)](#), we can obtain the function

$$f(z) = \exp(hz^2/2 + gz) - \exp(hz^2/2) - gu \tag{13}$$

and its first derivative

$$f'(z) = (hz + g) \exp(hz^2/2 + gz) - hz \exp(hz^2/2) \tag{14}$$

The corresponding reliability function is:

$$R(u) = 1 - F(u) = 1 - \Phi(z) \tag{15}$$

The pdf of U is expressed as:

$$f(u) = \varphi_Z \left[q_{g,h}^{-1}(u) \right] / q'_{g,h} \left[q_{g,h}^{-1}(u) \right] \tag{16}$$

and the log-type (Xu and Genton, 2015) is:

$$\begin{aligned} \phi(z) &= \log f(u) = \log \varphi(z) - \log q'_{g,h}(z) \\ &= -(1+h)z^2/2 - \log \left\{ \exp(gz) + g^{-1} [\exp(gz) - 1] hz \right\} - \log(2\pi)/2 \end{aligned} \tag{17}$$

The influence of g and h on the pdf and the reliability are illustrated in Figures 1(a) and 1(b). As g increases, the distribution will become more skewed. When $g=0$, U has a symmetric distribution. As h increases, the tails of the distribution get heavier. Moreover, all curves of reliability meet at the same point $F_U(0) = \Phi(0) = 0.5$, which is the median value.

2.3 Quantile-based estimation

Introducing the parameters of location $\mu \in R$ and scale $\sigma > 0$, the regression model is obtained as follows:

$$y = \mu + \sigma q_{g,h}(z) = \mu + \sigma g^{-1} [\exp(gz) - 1] \exp(hz^2/2) \tag{18}$$

Quantile-based parameter estimation approach was proposed by Tukey (1977), which used a letter value based sequence of quantiles to estimate the parameters sequentially. In this work, a brief description of the letter value based method is carried out with following steps (Hoaglin *et al.*, 1985; Rayner and Macgillivray, 2002b; Bee and Trapin, 2016):

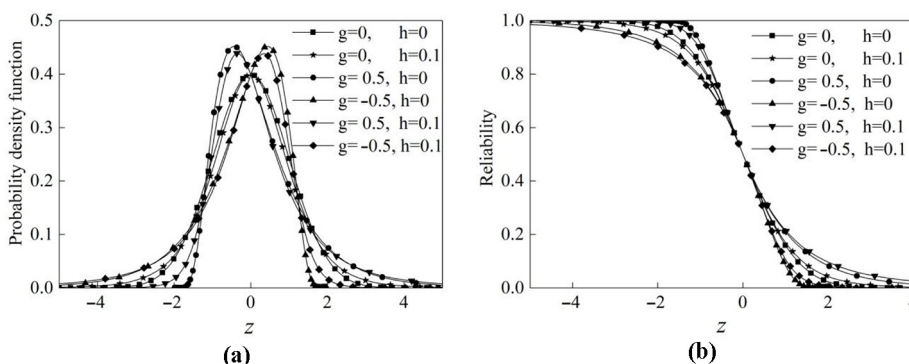


Figure 1.
Properties of Tukey's
g-and-h distribution

Notes: (a) Influence of g and h on the pdf; (b) influence of g and h on the reliability

- Estimate parameter μ by the median of the samples;
- For $z_p = -z_{1-p}$, then by

$$\frac{y_{1-p} - y_{0.5}}{y_{0.5} - y_p} = \frac{\left\{ \mu + g^{-1} \sigma [\exp(-gz_p) - 1] \exp\left(\frac{hz_p^2}{2}\right) \right\} - \mu}{\mu - \left\{ g^{-1} \sigma [\exp(gz_p) - 1] \exp\left(\frac{hz_p^2}{2}\right) + \mu \right\}} = \exp(-gz_p) \quad (19)$$

we can get:

$$g_p = -\frac{1}{z_p} \log\left(\frac{y_{1-p} - y_{0.5}}{y_{0.5} - y_p}\right) \quad (20)$$

Calculate $g_{p_i} = -\log[(y_{1-p_i} - y_{0.5})/(y_{0.5} - y_{p_i})]/z_{p_i}$ with median rank estimator p_i of samples, where y_{p_i} is the corresponding i -th sample quantile. Finally, the value g will be obtained by the mean of all g_{p_i} . In sequence, by

$$y_{1-p} - y_{0.5} = g^{-1} \sigma [\exp(-gz_p) - 1] \exp\left(\frac{hz_p^2}{2}\right) \quad (21)$$

the relationship is induced as:

$$UHS_p^* = g(y_{1-p} - y_{0.5}) [\exp(-gz_p) - 1]^{-1} = \sigma \exp\left(\frac{hz_p^2}{2}\right) \quad (22)$$

A plot of $\log(UHS_p^*)$ against $z_p^2/2$ would have intercept $\log(\sigma)$ and slope h .

- Regress $\log[(y_{1-p} - y_{0.5})/(\exp(-gz_p)-1)]$ or $\log[(y_{0.5} - y_{1-p})/(1 - \exp(-gz_p))]$ on $z_p^2/2$ depending on g positive or negative to estimate σ and h . By adjusting quantile selections, a quantile estimation method that works well for g and h estimation is demonstrated.

3. Parameter estimation of the g-and-h distribution

3.1 Degradation data description

Degradation data are generally obtained from repeated measurements with designed time points and certain number of units. The measurements are stopped until the degradation measurements reach the pre-specified failure threshold or the terminated time. The assumptions about the manner which the test and measurement should be conducted (Lu and Meeker, 1993) are as follows:

- Sample units are randomly selected from a population or production process and random measurement errors are independent across time.
- Sample units are tested in a particular homogenous environment such as the same constant temperature.
- Measurement (or inspection) times are pre-specified, the same across all the test assets, and may or may not be equally spaced in time.
- The lifetimes of the sample units are assumed to be independent and identically distributed random variables.

In Figure 2, the degradation data may be censored at the threshold, however, they are of great significance on the regression of the mean with time-dependent (Zhou *et al.*, 2012). So the rank of the censored data should be retained, and the degradation data should be ordered at every time point to make sure the rank. Consider that n units are tested, and the degradation measurements of each unit are available at time t_1, t_2, \dots, t_{m_i} . Given the degradation data (y_{ij}, t_j) , where y_{ij} is the measured deterioration for the i -th unit at time t_j , and $i = 1, 2, \dots, n, j = 1, 2, \dots, m_i$. The sample path with equation (18) considering time can be rewritten as:

$$\hat{y}_{ij} = \mu(t_j) + \sigma(t_j)u_{p_i} \tag{23}$$

where

$$u_{p_i} = q_{g,h}(z_{p_i}) = g^{-1}[\exp(gz_{p_i}) - 1] \exp(hz_{p_i}^2/2) \tag{24}$$

is the quantile expression of equation (1), and $z_{p_i} = \Phi^{-1}(p_i)$, with the median rank estimator:

$$\hat{p}_i = (i - 0.5)/n \tag{25}$$

Finally, with equations (8) and (9), the mean and the standard deviation (Std.) of Y can be obtained as:

$$\mu_Y = \mu(t) + \sigma(t)\mu_{g,h} \tag{26}$$

$$\sigma_Y = \sigma(t)\sigma_{g,h} \tag{27}$$

The location $\mu(t)$ and the scale $\sigma(t)$ can be selected as follows:

Linear regression:

$$a_0 + a_1t \tag{28}$$

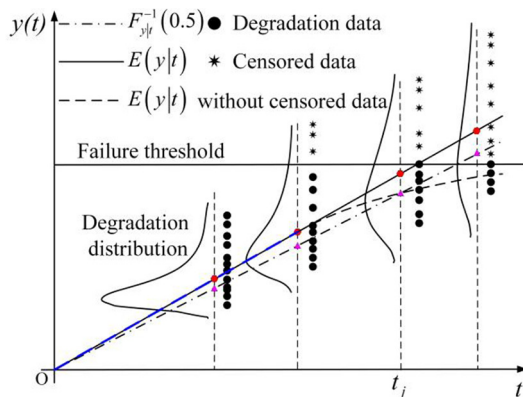


Figure 2. Description of degradation data

Polynomial regression:

$$a_0 + a_1t + a_2t^2 \tag{29}$$

Exponential regression:

$$a_0 \exp(a_1t + a_2t^2) \tag{30}$$

The reliability function evaluated at the time t is just the probability that the failure lifetime T beyond the time t :

$$R(t) = P\{T > t\} = 1 - F_T(t) \tag{31}$$

where $F(\cdot)$ denotes the cdf of the lifetime. For the increasing degradation path, the relationship between them is defined as:

$$P\{T > t\} = P\{y < D\} \tag{32}$$

where D is the failure threshold. Then the reliability can be obtained by [equation \(15\)](#):

$$R(t) = F(D) = 1 - \Phi \left[q_{g,h}^{-1} \left(\frac{D - \mu(t)}{\sigma(t)} \right) \right] \tag{33}$$

and solution of the reliability with [equation \(12\)](#).

According to the section 2.3, the quantile-based estimator is performed without considering the influence of time, which will lead the estimator to be complicated or inappropriate for the degradation modeling. Inspired by the quantile estimator ([Lu and Meeker, 1993](#); [Xu et al., 2014](#); [Takeuchi et al., 2003](#)), the unknown parameters will be estimated by minimizing the root-mean-square error (RMSE) on the degradation path of each unit as illustrated in [Figure 2](#) using the [equation \(34\)](#):

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n \left[\frac{1}{m_i} \sum_{j=1}^{m_i} (y_{ij} - \hat{y}_{ij})^2 \right]} \tag{34}$$

The rank of the censored data should be considered for computation. For convenient solution of the RMSE with many parameters, PSO may be a good choice for its good ability of global optimization and easy implement. It can perform parallel computation and simultaneously estimate all the parameters with the object function.

3.2 Particle swarm optimization

PSO ([Kennedy and Eberhart, 1995](#)) is an evolution technology based on the swarm intelligence, rooted from the research of bird predatory behavior. In the target searching space of N dimensions, there will be a group composed with M particles. The position vector of particle P_i in the t -th iteration is $\mathbf{X}_i(t) = (x_{i1}, x_{i2}, \dots, x_{iN})^T$ and the velocity vector $\mathbf{V}_i(t) = (v_{i1}, v_{i2}, \dots, v_{iN})^T$. The individual extreme point $pbest_i = (pbest_{i1}, pbest_{i2}, \dots, pbest_{iN})$ represents the influence of the memory of particles themselves and makes the particles have the ability of

global searching, meanwhile avoids the local minimum. The global extreme point $gbest_i = (gbest_{i1}, gbest_{i2}, \dots, gbest_{in})$ represents the influence of the group information, and reflects the information sharing as well as cooperation among the particles.

In the $(t + 1)$ -th iteration, the position and velocity of the particle P_i can be updated by the following equations:

$$v_{in}(t + 1) = \omega v_{in}(t) + c_1 rand() (pbest_{in} - x_{in}(t)) + c_2 rand() (gbest_{in} - x_{in}(t)) \quad (35)$$

$$x_{in}(t + 1) = x_{in}(t) + v_{in}(t + 1), i = 1, 2, \dots, M \quad (36)$$

where c_1 and c_2 are learning factors, and chosen to be 2 generally; $rand()$ is the random number between 0 and 1; ω is the inertia weight, which makes the method have better global search ability when big while tend to local search when small. In the computation, the expression of ω is:

$$\omega = \omega_{\max} - (\omega_{\max} - \omega_{\min}) \times Iter / Iter_{\max} \quad (37)$$

where $\omega_{\max} = 0.9$, $\omega_{\min} = 0.4$ (Shi and Eberhart, 1997); $Iter$ is the iterative number and $Iter_{\max}$ the maximum iterative number. To prevent the particles leaving the searching space, every dimensional velocity v_n of particles is confined in $[-v_{n\max}, v_{n\max}]$. That is to say, if $v_n > v_{n\max}$, then $v_n = v_{n\max}$; while $v_n = v_{n\max}$, then $v_n = -v_{n\max}$.

3.3 Procedure for parameter estimation with particle swarm optimization

- (1) Set the population size M , and specify the convergence precision and the maximum iterative steps;
- (2) Initialize the velocity and the position of the particles; and
- (3) Set the learning factors c_1 and c_2 , as well as the inertia weight ω ;
- (4) Give the fitness function namely the RMSE in this paper and the dimensions of the optimized parameters including g , h and the parameters in the location $\mu(t)$ and the scale $\sigma(t)$;
- (5) Minimum the fitness function and compute the fitness values of all particles;
- (6) Obtain the global extreme value and the individual extreme value, and update the velocity and the position of the particles;
- (7) Check whether the terminal condition is satisfied, if not, continue steps (5) to (6); if is, stop computation and output the optimum values.

3.4 Simulation for verification of the estimation algorithm

To verify the effectiveness of PSO in estimating parameters of the proposed approach based on Turkey's g -and- h distribution, a simulation is performed. Degradation measures are generated with the following at m equispaced time points for n units under the following model:

$$y_{ij} = (a_0 + a_1 t_j) + (b_0 + b_1 t_j) g^{-1} [\exp(gz_i) - 1] \exp(hz_i^2 / 2) \quad (38)$$

where $i = 1, 2, \dots, n; j = 1, 2, \dots, m, z = \xi + \varepsilon, \xi \sim N(0, 1), \varepsilon \sim N(0, 0.05)$. Different values of concerning parameters are listed in Table I. Considering the influence of the sample number and verification of different parameters, four groups of simulations with ten times are carried out. To control the big fluctuation at different time points, only one group of simulation data obeying the standard normal distribution is generated at every time, but a small disturb belonging to the standard normal distribution with mean 0, variance 0.05 is added at fifteen time points as illustration in Figure 2. In PSO, 1000 particles are set for computation. Mean and variance of the estimated parameters are computed, as well the bias between the true value and the mean. The results have proved that the approach could satisfy requirement of the estimation accuracy for the following study in degradation reliability investigation.

4. Applications

4.1 Degradation description

4.1.1 GaAs laser device. In this section, a practical case related to the degradation dataset of GaAs laser devices is presented to demonstrate the validity of the proposed estimation method (Meeker and Escobar, 1998). The operation current is measured as the degradation characteristic at 80°C, which increases over time. The device fails when the per cent increase in operating current exceeds a pre-specified threshold $D = 10$. Degradation data from 15 testing samples are collected, with observations made at assigned time point and plotted in Figure 3(a). It has the degradation trend of linearity, and expressions of the location and the scale by equation (28) respectively are:

$$\mu(t) = a_0 + a_1t \tag{39}$$

$$\sigma(t) = b_0 + b_1t \tag{40}$$

4.1.2 Crack growth. Sets of fatigue crack growth data versus time (in million cycles) are plotted in Figure 3(b). These data sets were collected on twenty-one test specimens. All specimens have an initial crack length of 0.90 inch. The failure threshold is set as a critical

Group	Sample number	Time points	Statistics	a_0	a_1	b_0	b_1	g	h
1	15	15	Mean	0.101	0.401	0.104	0.100	0.404	0.141
			Variance	0.002	0.002	0.001	0.001	0.053	0.027
			True values	0.1	0.4	0.1	0.1	0.4	0.1
			Bias	0.001	0.001	0.004	0.000	0.004	0.041
2	20	15	Mean	0.108	0.407	0.106	0.107	0.433	0.095
			Variance	0.001	0.001	0.001	0.001	0.055	0.012
			True values	0.1	0.4	0.1	0.1	0.4	0.1
			Bias	0.008	0.007	0.006	0.007	0.033	-0.005
3	15	15	Mean	0.202	0.599	0.046	0.291	0.283	0.180
			Variance	0.000	0.010	0.000	0.003	0.111	0.057
			True values	0.2	0.6	0.05	0.3	0.3	0.2
			Bias	0.002	-0.001	-0.004	-0.009	-0.017	-0.02
4	15	15	Mean	0.202	0.603	0.041	0.247	0.249	0.095
			Variance	0.000	0.008	0.000	0.005	0.073	0.021
			True values	0.2	0.6	0.05	0.3	0.2	0.1
			Bias	0.002	0.003	-0.009	-0.053	0.049	-0.005

Table I.
Parameter estimation

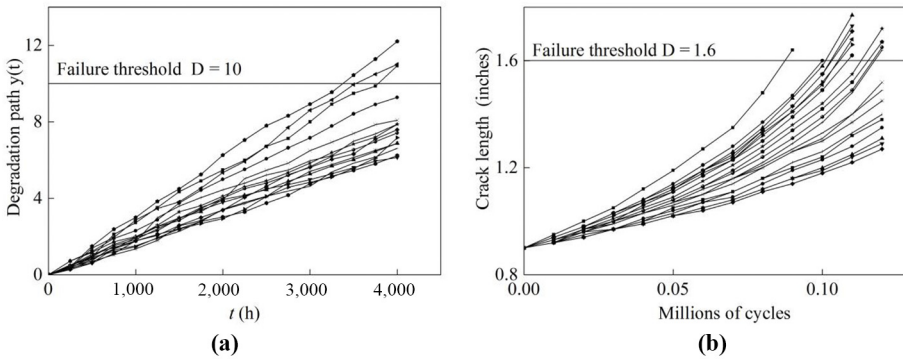


Figure 3.
Laser degradation data

Notes: (a) GaAs laser device; (b) crack growth

crack length of 1.6 inches. It is noticed that about half of the units do not fail by the end of the test. Obviously, a polynomial function by equation (29) will be necessary to fit the location and the scale with time:

$$\mu(t) = a_0 + a_1t + a_2t^2 \tag{41}$$

$$\sigma(t) = b_0 + b_1t + a_2t^2 \tag{42}$$

Minimizing the RMSE by PSO, the parameters could be estimated. First, the two important variables mean and Std. are computed by equations (26) and (27), and verified by comparison with the experiment data. Second, the proposed approach is applied in the lower thresholds where no failure occurs. The pseudo-lifetime data obtained by the interpolation in the lower thresholds are used to verify the reliability estimation. Then the proposed approach is used for the practical threshold and compared with others' results. The two-stage verification can guarantee the validity of the approach, and the whole procedure is illustrated in Figure 4.

4.2 Parameter estimation

The particle number in PSO is set 300, convergence precision 1-e20. For GaAs laser device, the time is divided 250 for scale match, and the iteration is stopped until 1593 steps. For the crack growth, the iteration is stopped until 1740 steps. Parameters of Tukey's g-and-h distribution estimated with PSO are listed in Table II and the RMSE versus iterative numbers is plotted in Figure 5. The mean and the variance at every test time are computed by equations (26) and (27) and compared with the experiment data in Figure 6, while the relative errors of the mean and the Std. are computed by equation (43) expressed as:

$$\delta = 100\% \times \left| \frac{\zeta - \hat{\zeta}}{\zeta} \right| \tag{43}$$

where ζ represents the experiment and $\hat{\zeta}$ the simulation, and the results show that they both match well.

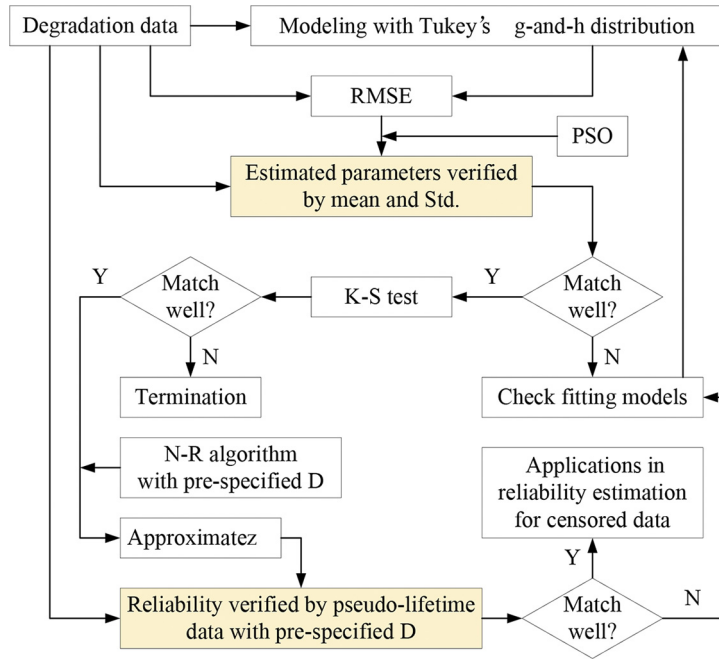
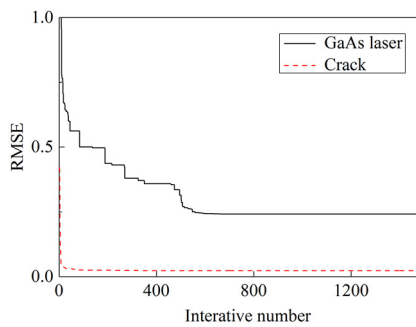


Figure 4. Illustration of the procedure for reliability estimation

Parameter	a_0	a_1	a_2	b_0	b_1	b_2	g	h	RMSE
GaAs laser	0.0051	0.4886	–	0.035	0.1006	–	0.4568	0.0	0.2411
Crack	0.9128	1.5147	33.7608	0.0050	0.0	15.5443	0.1723	0.0	0.0222

Table II. Parameter estimation

Figure 5. RMSE with iterative numbers by PSO



4.3 Reliability

KS-test is carried out for normal, Weibull and the g-h distribution. The g-h distribution fails to pass at time 250 for GaAs laser device, as well as at 0.01, 0.02 and 0.03 millions of cycles for crack growth, as shown in Figure 7. The reason may be that the data at the beginning are too concentrated. However, the g-h distribution shows better fitness than the other two

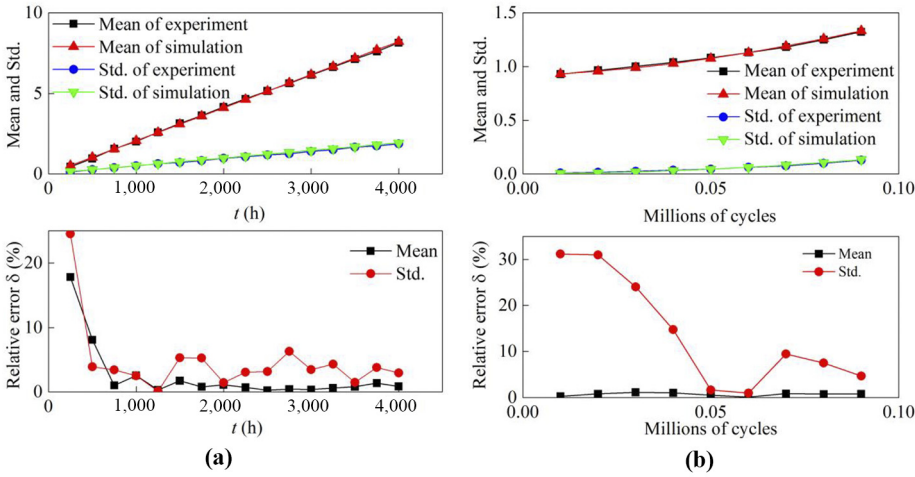


Figure 6. Simulation of mean and std. with relative error

Notes: (a) GaAs laser device; (b) crack growth

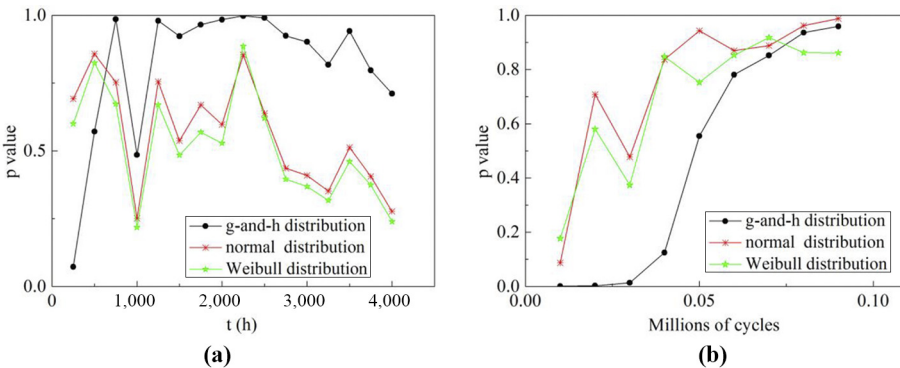


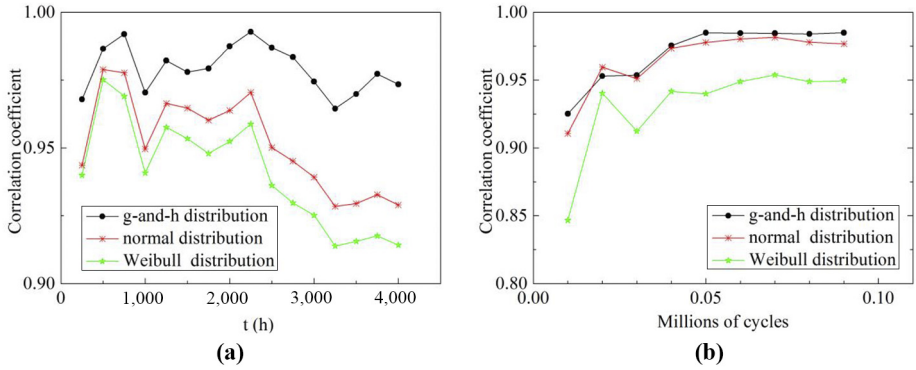
Figure 7. K-S test of normal, Weibull and the g-h distribution

Notes: (a) GaAs laser device; (b) crack growth

distributions at the rest time for the two practical applications with the correlation coefficients in Figure 8. In further, the Q-Q plots at the time $t = 4000$ h of the GaAs laser device and at 0.09 million cycles of the crack growth are respectively extracted as shown in Figure 9, which shows the advantage of the g-h distribution in fitting data distributions. Figure 10 gives the pdf of the degradation data varying with time for GaAs laser device and crack growth respectively. By equation (33), the degradation reliability is computed.

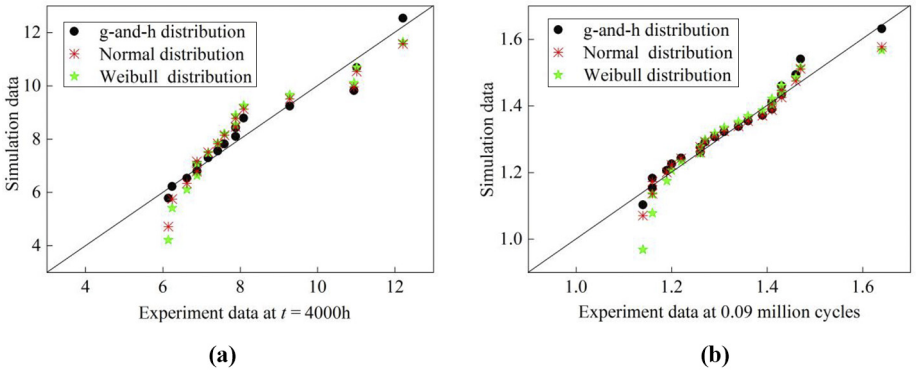
To verify the effectiveness of the approach for degradation reliability, the pseudo-lifetime data are obtained by interpolation at three different thresholds $D = 4, D = 5, D = 6$ of the GaAs laser device and $D = 1.2, D = 1.3, D = 1.4$ of the crack growth without censored data in Figure 11. Finally, the practical reliability for the GaAs laser device at failure threshold $D = 10$ and the crack at failure threshold $D = 1.6$ are computed respectively and compared with others' research results as shown in Figure 12, where the pseudo lifetime are obtained by extrapolating the censored data to the failure threshold, especially the first two lifetime data of the GaAs laser device and the first 12 lifetime data of the crack growth are practically failed.

Figure 8.
Correlation coefficients of normal, Weibull and the g-h distribution



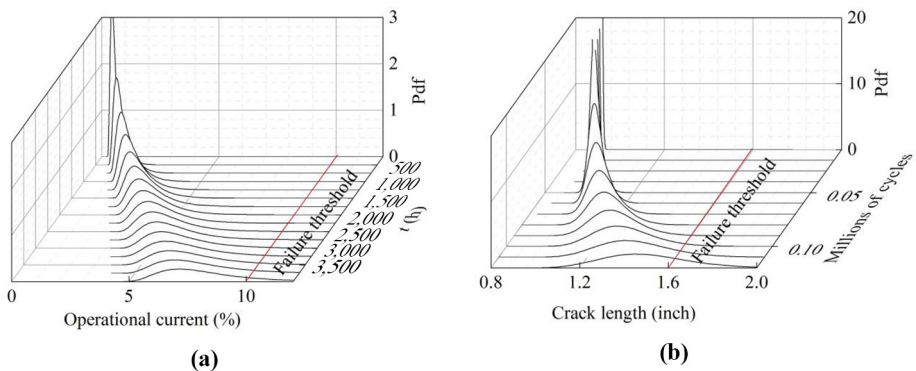
Notes: (a) GaAs laser device; (b) crack growth

Figure 9.
Q-Q plot of normal, Weibull and the g-h distribution



Notes: (a) GaAs laser device; (b) crack growth

Figure 10.
Pdf over time



Notes: (a) GaAs laser device; (b) crack growth

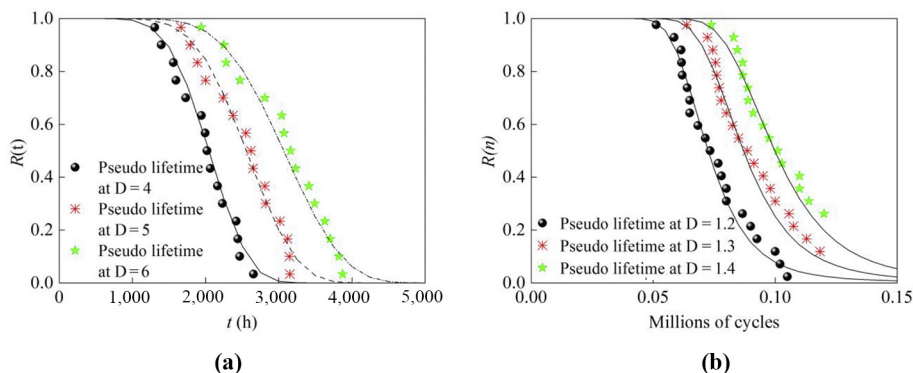


Figure 11. Verification by lower failure threshold

Notes: (a) GaAs laser device; (b) crack growth

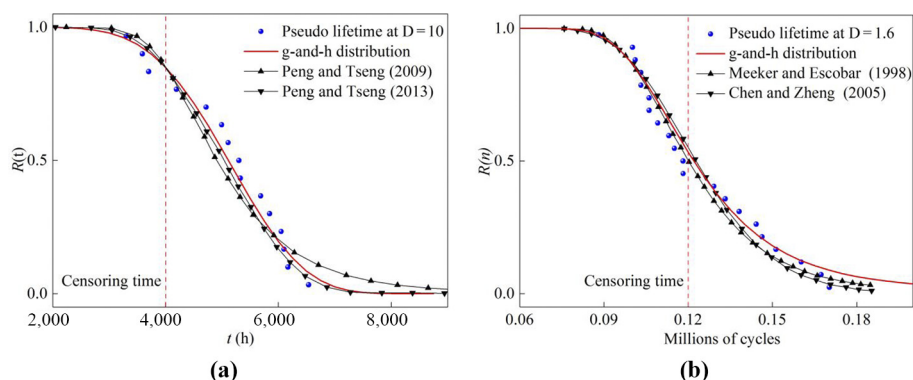


Figure 12. Comparison with others' work

4.4 Results

Regression of the degradation paths with Tukey's g-and-h distribution shows good fitness with the experiment data. Comparisons of the mean and the Std. between the simulation and the experiment data prove that PSO with RMSE for parameter estimation is effective. The approach with Tukey's g-and-h distribution for reliability estimation can well match the practical lifetime point at the beginning of failure, which is concerned with the tail the skewness of the distribution. Compared with other methods, the Tukey's g-and-h distribution for degradation reliability could have a good fit of the pseudo-lifetime at the descent stage of the curve as well.

From the above analysis, the skewness and the tail of the distribution should be considered to obtain reasonable results when modeling the degradation processes, because they can exactly predict the beginning time of the failure. Although the proposed approach may introduce more parameters, they make the model more flexible and reliable results can be governed. However, an important parameter h has to be restricted in the interval $[0, \infty]$ when $h < 0$, in order to approximate the distribution of the degradation data.

5. Conclusions

In this paper, an approach based on Tukey's g-and-h distribution for degradation reliability is proposed. PSO as an intelligent optimization algorithm is investigated for the parameter estimation and verified with the simulation study. Aim at no explicit expression of Tukey's g-and-h distribution, the Newton-Raphson algorithm is adopted for solution of reliability. The effectiveness of the proposed approach for the reliability estimation have been demonstrated through two practical degradation data. Clearly, the approach can well estimate the reliability by the two-stage verification. In further, it can well predict the reliability with censored data compared with others' results.

References

- Badrinath, S.G. and Chatterjee, S. (1991), "A data-analytic look at skewness and elongation in common-stock-return distributions", *Journal of Business and Economic Statistics*, Vol. 9 No. 2, pp. 223-233.
- Bee, M. and Trapin, L. (2016), "A simple approach to the estimation of Tukey's gh distribution", *Journal of Statistical Computation and Simulation*, Vol. 86 No. 16, pp. 3287-3302.
- Chaudhuri, A. and Ghosh, S.K. (2016), *Quantitative Modeling of Operational Risk in Finance and Banking Using Possibility Theory*, Springer, Basel.
- Chen, D. and Zhao, C. (2009), "Data-driven fuzzy clustering based on maximum entropy principle and PSO", *Expert Systems with Applications*, Vol. 36 No. 1, pp. 625-633.
- Degen, M., Embrechts, P. and Lambrigger, D.D. (2007), "The quantitative modeling of operational risk: between g-and-h and EVT", *ASTIN Bulletin*, Vol. 37 No. 2, pp. 265-291.
- Dupuis, D.J. and Field, C.A. (2004), "Large wind speeds: modeling and outlier detection", *Journal of Agricultural, Biological, and Environmental Statistics*, Vol. 9 No. 1, pp. 105-121.
- Fischer, M. (2010), "Generalized Tukey-type distributions with application to financial and teletraffic data", *Statistical Papers*, Vol. 51 No. 1, pp. 41-56.
- Gorjian, N., Ma, L., Mittinty, M., Yarlagadda, P. and Sun, Y. (2009), "A review on degradation models in reliability analysis", *Proceedings of the 4th World Conference on Engineering Asset Management*, pp. 369-384.
- He, Y. and Raghunathan, T.E. (2006), "Tukey's gh distribution for multiple imputation", *The American Statistician*, Vol. 60 No. 3, pp. 251-256.
- Headrick, T.C., Kowalchuk, R.K. and Sheng, Y. (2008), "Parametric probability densities and distribution functions for Tukey g-and-h transformations and their use for fitting data", *Applied Mathematical Sciences*, Vol. 2 No. 9, pp. 449-462.
- Hoaglin, D.C., Mosteller, F. and Tukey, J.W. (1985), *Exploring Data Tables, Trends, and Shapes*, Wiley, New York, NY.
- Jiménez, J.A. and Arunachalam, V. (2011), "Using Tukey's g and h family of distributions to calculate value-at-risk and conditional value-at-risk", *The Journal of Risk*, Vol. 13 No. 4, pp. 95-116.
- Jorge, M. and Boris, I. (1984), "Some properties of the Tukey g and h family of distributions", *Communication in Statistics-Theory and Methods*, Vol. 13 No. 3, pp. 353-369.
- Kennedy, J. and Eberhart, R. (1995), "Particle swarm optimization", *Proceedings of IEEE International Conference on Neural Network*, Vol. 4, pp. 1942-1948.
- Klein, I. and Mittnik, S. (2002), *Contributions to Modern Econometrics from Data Analysis to Economic Policy*, Springer Science+Business Media, New York, NY.
- Lu, J.C. and Meeker, W.Q. (1993), "Using degradation measures to estimate a time-to-failure distribution", *Technometrics*, Vol. 35 No. 2, pp. 161-174.
- Macgillivray, H.L. (1992), "Shape properties of the g-and-h and Johnson families", *Communications in Statistics-Theory and Methods*, Vol. 21 No. 5, pp. 1233-1250.

-
- Majumder, M.M.A. and Ali, M.M. (2008), "A comparison of methods of estimation of parameters of Tukey's gh family of distributions", *Pakistan Journal of Statistics*, Vol. 24 No. 2, pp. 135-144.
- Morgenthaler, S. and Tukey, J.W. (2000), "Fitting quantiles: doubling, HR, HQ, and HHH distributions", *Journal of Computational and Graphical Statistics*, Vol. 9 No. 1, pp. 180-195.
- Meeker, W.Q. and Escobar, L.A. (1998), *Statistical Methods for Reliability Data*, John Wiley and Sons, New York, NY.
- Örkcü, H.H., Özsoy, V.S., Aksoy, E. and Dogan, M.I. (2015), "Estimating the parameters of 3-p, Weibull distribution using particle swarm optimization: a comprehensive experimental comparison", *Applied Mathematics and Computation*, Vol. 268 No. 9, pp. 201-226.
- Rayner, G.D. and Macgillivray, H.L. (2002a), "Numerical maximum likelihood estimation for the g-and-k and generalized g-and-h distributions", *Statistics and Computing*, Vol. 12 No. 1, pp. 57-75.
- Rayner, G.D. and Macgillivray, H.L. (2002b), "Weighted quantile-based estimation for a class of transformation distributions", *Computational Statistics and Data Analysis*, Vol. 39 No. 4, pp. 401-433.
- Shi, Y. and Eberhart, R. (1997), "A modified particle swarm optimizer", *IEEE International Conference on Evolutionary Computation*, Vol. 6, pp. 69-73.
- Takeuchi, I., Yamanaka, N. and Furuhashi, T. (2003), "Robust regression under asymmetric or/and non-constant variance error by simultaneously training conditional quantiles", *Proceedings of the International Joint Conference on Neural Networks*, Vol. 3, pp. 1729-1734.
- Tukey, J.W. (1977), "Modern techniques in data analysis", *NSF-Sponsored Regional Research Conference*.
- Xu, G. and Genton, M.G. (2015), "Efficient maximum approximated likelihood inference for Tukey's g-and-h distribution", *Computational Statistics and Data Analysis*, Vol. 91, pp. 78-91.
- Xu, Y., Iglewicz, B. and Chervoneva, I. (2014), "Robust estimation of the parameters of g-and-h distributions, with applications to outlier detection", *Computational Statistics and Data Analysis*, Vol. 75 No. 7, pp. 66-80.
- Ye, Z.S. and Xie, M. (2015), "Stochastic modelling and analysis of degradation for highly reliable products", *Applied Stochastic Models in Business and Industry*, Vol. 31 No. 1, pp. 16-32.
- Zhou, R., Gebraeel, N. and Serban, N. (2012), "Degradation modeling and monitoring of truncated degradation signals", *IIE Transactions*, Vol. 44 No. 9, pp. 793-803.
- Zuo, M.J., Jiang, R. and Yam, R.C.M. (1999), "Approaches for reliability modeling of continuous-state devices", *IEEE Transactions on Reliability*, Vol. 48 No. 1, pp. 9-18.

Further reading

- Chen, Z. and Zheng, S. (2005), "Lifetime distribution based degradation analysis", *IEEE Transactions on Reliability*, Vol. 54 No. 1, pp. 3-10.
- Peng, C.Y. and Tseng, S.T. (2009), "Mis-specification analysis of linear degradation model", *IEEE Transactions on Reliability*, Vol. 58 No. 3, pp. 444-455.
- Peng, C.Y. and Tseng, S.T. (2013), "Statistical lifetime inference with skew-Wiener linear degradation models", *IEEE Transactions on Reliability*, Vol. 62 No. 2, pp. 338-350.

Corresponding author

Yongqiang Zhao can be contacted at: 13B308008@hit.edu.cn

For instructions on how to order reprints of this article, please visit our website:

www.emeraldgroupublishing.com/licensing/reprints.htm

Or contact us for further details: permissions@emeraldinsight.com