IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received April 21, 2019, accepted May 2, 2019, date of publication May 7, 2019, date of current version May 21, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2915279

Nonlinear Least Squares Estimation for Parameters of Mixed Weibull Distributions by Using Particle Swarm Optimization

ZHONG LU[1], LI DONG[1], AND JIA ZHOU[2]
[1]College of Civil Aviation, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China
[2]Department of Aircraft Maintenance, China Eastern Airlines Jiangsu Ltd., Nanjing 211106, China

Corresponding author: Zhong Lu (luzhong@nuaa.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant U1733124 and in part by the Fundamental Research Funds for the Central Universities of NUAA under Grant 3082018NT2018019.

ABSTRACT Mixed Weibull distributions are widely used in lifetime modeling of products with multiple failure modes. It is difficult to estimate parameters of the mixed Weibull distribution since it contains multiple parameters. A parameter estimation model for the mixed Weibull distributions is proposed based on nonlinear least squares estimation (LSE). An approach of determining parameters' approximate values and rough bounds is presented for selecting good starting points used in the particle swarm optimization (PSO) procedure. The PSO solution of the nonlinear LSE method is proposed by a step-by-step procedure. A case study is given to illustrate the accuracy and efficiency of our proposed method. Compared with the genetic algorithm (GA)-based nonlinear LSE method, our method shows advantages in both accuracy and efficiency. And compared with the maximum likelihood estimation (MLE) method, our method shows significant advantages in efficiency.

INDEX TERMS Mixed Weibull distribution, parameter estimation, nonlinear least squares estimation, particle swarm optimization.

I. INTRODUCTION

The Weibull distribution has been proved, empirically and analytically, to be an excellent model for many individual failure mechanisms on individual components. Its accuracy in modeling single failure modes has been widely recognized. On the other hand, a Weibull distribution is often a poor model for a system or product with multiple failure modes. There are times when a failure distribution for a product with multiple failure modes needs to be determined. A preferred approach is to perform a Weibull analysis on each mode separately if enough data is available, and then combine them mathematically to obtain the overall product distribution. However, there are usually few failures on the individual modes. What's worse, product failures are reported but details of the failure modes are not. In this case, a Mixed Weibull distribution may be a good solution [1].

The probability density function (PDF) of products with multiple failure modes will have a bimodal or multimodal shape. Consequently, the mixed Weibull distributions can yield the best fits to lifetimes of such products and have been widely used in this field. He *et al.* [2] presented a mixed Weibull probability distribution model for reliability evaluation of the paper-oil insulation. Ling *et al.* [3] studied the fatigue life prediction of mechanical components based on the mixed Weibull distribution, and a case study is used to show that their method can fit the observed cumulative failure probability more properly than the standard two-parameter Weibull distribution. Reliability predictions are used to illustrate the effectiveness of their model. Carlucci and Tognarelli [4] presented an approach to model the forced outage distribution by using the mixed Weibull Distribution. This approach can be applied to support reliability block diagram based Monte Carlo simulation. Du *et al.* [5] established a reliability model for electrical engineering products by using the mixed Weibull distribution. The results indicate that the mixed Weibull distribution model has a higher accuracy compared with other methods. Andersen and Dennison [6] developed a mixed Weibull Distribution Model of DC Dielectric Breakdown failures, as these failures incorporate both low-and high-energy defect modes. Ghavijorbozeh and Hamadani [7] applied the mixed Weibull

The associate editor coordinating the review of this manuscript and approving it for publication was Kan Liu.

distribution in the machine reliability analysis for the Cellular manufacturing system. Their findings show that reliability analysis model based on the mixed Weibull distribution approach can give options to a user to select the suitable failure rate and mode for a specific situation. Yuan *et al.* [8] proposed a mixed Weibull distribution based reliability estimation method for aircraft engines, which are typical complex products with diverse failure modes.

Although the mixed Weibull distribution is preferred to be applied in modeling of products with multiple failure modes, its parameters cannot be estimated easily. Many researchers have been making efforts to get a good solution of the parameter estimators for the mixed Weibull distribution. Jiang and Kececioglu [9] proposed an algorithm in terms of the principle of the Maximum Likelihood Estimation (MLE) via the Expectation and Maximization algorithm, and it is derived for both postmortem and non-postmortem time-to-failure data. Attardi *et al.* [10] proposed an MLE algorithm of the mixed Weibull model parameters. And a stepwise procedure has been used to test the significance of covariates and to construct the regression model. Gong [11] used the shuffled complex-evolution metropolis algorithm to estimate mixed Weibull distribution parameters in the automotive reliability analysis. And the accuracy of the algorithm is verified by comparing with the MLE method. Huang *et al.* [12] proposed an algorithm to estimate parameters for mixed Weibull distributions by combining the fuzzy classification and the MLE. The results show their algorithm has better accuracy compared with Jiang and Kececioglu's method [9]. Touw [13] studied the Bayesian estimation of mixed Weibull distributions. And it is found that the Bayesian estimation can improve the accuracy for situations with low mixture ratios, so long as the prior on the weak subpopulation's characteristic life has an expected value less than or equal to the true characteristic life. Ling *et al.* [14] establishes the parameter estimation methods for the mixed Weibull model using the theory of the nonlinear least squares estimation (LSE); and quasi-Newton method is used to solve the optimization problem. And, a numerical example is given to compare the proposed method with the conventional graphical method. Krohling *et al.* [15] formulated the parameter estimation of the mixed Weibull distribution as the maximization of the likelihood function of the corresponding mixture model, and Bare Bones Particle Swarm Optimization (PSO) algorithm were applied to get the maximum value of the likelihood function. Chi *et al.* [16], [17] presented an optimization model for parameter estimations for the mixed Weibull distribution by minimizing the residual sum of squares, and the Cuckoo Search algorithm is used to search the optimal estimators. At present, the MLE and the LSE are the most frequently used methods for distribution parameter estimations. MLE is considered to have many good statistical properties and is preferred by researchers. The LSE method is essentially a regression method, which takes the summation of squared residuals as the minimizing objective. LSE is very convenient, and hence is preferred by practitioners [18].

Among all of the optimization algorithms, PSO is a population based stochastic optimization technique, which mimics the behavior of flocks of birds, swarms of insects or schools of fish, in which individuals are called particles and the population is called a swarm [19]. The PSO algorithm has been widely used in parameter estimation since it has been proposed by Eberhart and Kennedy more than twenty years ago. özsoy *et al.* [20] applied the PSO algorithm in the parameter estimation based on MLE for the four-parameter Burr type III distribution, and their simulation shows that the PSO approach provides accurate estimates. Örkcü *et al.* [21] also presented a comprehensive investigation of different PSO variants (according to inertia weight procedures, acceleration coefficients, particle size, and search space) in the parameter estimation of the three-parameter Weibull distribution, and numerical examples show that PSO approach variants exhibit a high efficiency and accuracy in the MLE based parameter estimation. Jiang *et al.* [22] conducted the parameter estimation for Weibull, Rayleigh, Gamma and Lognormal probability distributions by using the LSE method, the MLE method and the moment method respectively. Three metaheuristic optimization algorithms, including the bat algorithm, the cuckoo search algorithm and the PSO algorithm, as well as the numerical method are employed to obtain the optimal parameters. Their experimental results show that the three metaheuristic optimization algorithms are better than the numerical method. Additionally, the PSO based estimation for the Weibull distribution is slightly superior to the cuckoo search based method. Lu *et al.* [23] proposed the nonlinear least-squares parameter evaluation methods of maintenance time distribution based on PSO, their result shows the precision of method proposed here is better than the traditional methods. Ma *et al.* [24]–[26] studied the application of optimization method in the gas emission source identification. Different kinds of swarm intelligent optimization methods are discussed in their work, which include the PSO method, the genetic algorithm (GA), the simulated annealing, the ant colony optimization algorithm, the firefly algorithm, the pattern search method, the Nelder-Mead simplex method and their hybrid optimization methods. The comparison among these methods has been presented and their results show the PSO method is a good choice in the gas emission source identification, and it can obtain satisfied estimation results under different boundary constraints.

PSO algorithm is a type of global search method, which depends less on the initial values compared with traditional optimization methods. PSO method is superior in computational efficiency compared with other metaheuristic algorithms such as the ant colony optimization algorithm, the firefly algorithm and the GA. In addition, PSO method can obtain satisfied estimation results under different boundary constraints [24], [25]. Therefore, it will also be applied in the LSE based parameter estimation of the mixed Weibull distribution in this study. Specifically, the PSO algorithm will be used to minimize the summation of the relative deviations

between the actual value of the distribution function and the value predicted by the distribution model.

The rest of this paper is structured as follows. In Section II, definitions related to the mixed Weibull distribution and the empirical distribution function are given, and the optimization model of the LSE based parameter estimation is developed. In Section III, the method of determining parameters' approximate values and rough bounds of the mixed Weibull distribution is given for selecting good starting points used in PSO, and the PSO solution of the LSE based optimization model for parameter estimations is proposed by a step by step procedure. In Section IV, case studies are given to illustrate the correctness and effectiveness of our proposed method. In Section V, concluding remarks are presented.

II. PROBLEM STATEMENT

A. DEFINITIONS RELATED TO MIXED WEIBULL DISTRIBUTIONS

The PDF of the mixed Weibull distribution is expressed as

$$f(t)=\sum_{i=1}^{n} p_i \frac{\beta_i}{\alpha_i^{\beta_i}} t^{\beta_i-1} e^{-\left(\frac{t}{\alpha_i}\right)^{\beta_i}}, \tag{1}$$

where p_i, β_i and α_i are the weighting factor, the shape parameter and the scale parameter of the ith subpopulation respectively, and n is the number of sunpopulations [18]. And we have

$$\sum_{i=1}^{n} p_i = 1. \tag{2}$$

The distribution function or the cumulative density function (CDF) of the mixed Weibull distribution is expressed as

$$F(t)=1-\sum_{i=1}^{n} p_i e^{-\left(\frac{t}{\alpha_i}\right)^{\beta_i}}. \tag{3}$$

In reliability engineering, CDF is also called the failure probability or unreliability, and the reliability function of Weibull distributions can be expressed as

$$R(t)=\sum_{i=1}^{n} p_i e^{-\left(\frac{t}{\alpha_i}\right)^{\beta_i}}. \tag{4}$$

A matrix used to denote the unknown parameters in (1), (3) or (4) is given as

$$\boldsymbol{\theta}=\begin{bmatrix} p_1 & p_2 & \dots & p_n \\ \beta_1 & \beta_2 & \dots & \beta_n \\ \alpha_1 & \alpha_2 & \dots & \alpha_n \end{bmatrix}, \tag{5}$$

where we have

$$\begin{cases} 0<p_i<1 \\ \sum_{i=1}^{n} p_i = 1 \\ \alpha_i>0 \\ \beta_i>0 \end{cases}, \quad i=1,2,\cdots,n. \tag{6}$$

B. EMPIRICAL DISTRIBUTION FUNCTIONS

In failure data analysis, the empirical distribution function is usually used to denote the actual value of the distribution function. To the m ordered lifetime samples $t_1,\ t_2,\ \cdots,\ t_m$, i.e. $t_1 \le t_2 \le \cdots \le t_m$, the empirical distribution function can be given as [27]

$$\hat{F}(t)=\begin{cases} 0, & t<t_1 \\ \dfrac{j}{m}, & t_j \le t < t_{j+1}\ (j=1,2,\cdots,m-1) \\ 1, & t \ge t_m. \end{cases} \tag{7}$$

When the size of lifetime samples are less than 20, the median rank values of $F(t)$ are computed using Bernard's approximation which is given below [28]

$$\hat{F}(t)=\begin{cases} 0, & t<t_1 \\ \dfrac{j-0.3}{m+0.4}, & t_j \le t < t_{j+1}\ (j=1,2,\cdots,m-1) \\ 1, & t \ge t_m. \end{cases} \tag{8}$$

C. OPTIMIZATION MODEL OF THE NONLINEAR LSE BASED PARAMETER ESTIMATION

The objective of parameter estimations is to find the parameter values of the model to best fit the data samples. The fit of the distribution is usually measured by the deviations between the actual value of the empirical distribution function and the value predicted by the distribution model. In the linear LSE, the CDF of a distribution model is generally transformed to a linear formation for the convenience of calculation. However, it is very difficult, if not impossible, to transform the CDF of the mixed Weibull distribution to a linear formation. In this study, the nonlinear CDF of the mixed Weibull distribution is used to calculate the residuals directly. And the summation of the squared residuals is selected as the objective function of the parameter estimation model. With the expression of constraints for the unknown parameters given in (6), the optimization model of the nonlinear LSE based parameter estimation can be expressed as

$$\min f(\boldsymbol{\theta})=\sum_{j=1}^{m}\left[\hat{F}(t_j)-F(t_j,\boldsymbol{\theta})\right]^2$$
$$\text{s.t.} \begin{cases} 0<p_i<1 \\ \sum_{i=1}^{n} p_i = 1 \\ \alpha_i>0 \\ \beta_i>0 \end{cases}, \quad i=1,2,\cdots,n, \tag{9}$$

where $\hat{F}(t_j)$ is the value of the empirical distribution function, and it can be calculated via (7) or (8). $F(t_j,\boldsymbol{\theta})$ is the CDF value that can be calculated via (3).

In (9), the residuals are expressed by the difference between the nonlinear CDF and the empirical distribution. In this way, we do not have to transform the CDF to a linear formation for using of the linear LSE method, and the difficulties in linearization of the mixed Weibull distributions' CDF can be overcomed.

Our goal is to find the estimating values of all parameters that can minimize the objective function in the case that all constraints can be satisfied. And the PSO algorithm will be used to get the optimizing values of these parameters in the following section.

III. PSO SOLUTION FOR THE LSE BASED PARAMETER ESTIMATION OF THE MIXED WEIBULL DISTRIBUTION

PSO is a metaheuristic optimization method. It solves an optimization problem by having a population (swarm) of candidate solutions (particles), and moving these particles around in the search-space according to a few simple formulae over the particle's position and velocity. The movements of the particles are guided by their own (local) best known position in the search-space as well as the entire swarm's (global) best known position. Better positions of the particles maybe discovered by the movement. The discovery of the better positions will update the local and global best known positions, and guide the movements of the swarm. The process is repeated, and by doing so, it is hoped that a satisfactory solution will eventually be discovered [29].

A. APPROACH OF DETERMINING PARAMETERS' APPROXIMATE VALUES AND ROUGH BOUNDS

PSO is a space searching optimization, the goodness of the starting point will affect the efficiency significantly. Approaches for determining parameters' approximate values and rough bounds are proposed here, and then the starting points (namely initial points) can be obtained according to the rough bounds. Although the method proposed can only get parameters' approximate values and rough bounds, it will not affect the accuracy of the estimating results. As the approximate values and rough bounds are only used to determine the initial points, and the accuracy of the PSO algorithm depends less on the initial points.

The empirical distribution function of a single Weibull distribution is expressed as

$$F(t) = 1 - e^{-\left(\frac{t}{\alpha}\right)^{\beta}}. \tag{10}$$

It can be transferred to the following linear formation

$$\ln\{-\ln[1 - F(t)]\} = \beta \ln t - \beta \ln \alpha. \tag{11}$$

To lifetime samples $t_j (j = 1, 2, \cdots, m)$ following the mixed Weibull distribution with n subpopulations, and their corresponding empirical distribution functions $\hat{F}(t_j)$, we do transformations according to (12)

$$\begin{cases} x_j = \ln t_j \\ y_j = \ln\left\{-\ln\left[1 - F(t_j)\right]\right\} \end{cases} \quad j = 1, 2, \cdots, m. \tag{12}$$

When plotting all the m points (x_j, y_j) on the general coordinate system, there will be n straight lines which are separated by $n - 1$ inflection points, and each straight line is roughly corresponding to a subpopulation of the mixed Weibull distribution.

By (11), the slope of the ith straight line is the approximate value of the shape parameter β_i.

Assume the size of points fitting the ith straight line is $m_i (i = 1, 2, \cdots, n)$, then the approximate values of the weighting factors can be denoted as

$$p_i = \frac{m_i}{m}. \tag{13}$$

According to (7) and (10), we have

$$\begin{cases} F(\alpha) = \dfrac{j^{(\alpha)}}{m} \\ F(\alpha) = 1 - e^{-1}, \end{cases} \tag{14}$$

where $j^{(\alpha)}$ denotes the ordinal number of the lifetime sample whose value is roughly equal to α. Hence we have

$$j^{(\alpha)} = \left[m \cdot (1 - e^{-1})\right]. \tag{15}$$

where $[\bullet]$ is the rounding function. For the ith subpopulation of the mixed Weibull distribution, we have the similar expression

$$j_i^{(\alpha)} = \left[m_i \cdot (1 - e^{-1})\right]. \tag{16}$$

Thus, the $j_i^{(\alpha)}$th lifetime sample pertinent to the ith straight line is the approximate value of the scale parameter α_i.

Based on the parameters' approximate values determined, we can determine the rough bound of each parameter. Assume $p_i^{(A)}$, $\alpha_i^{(A)}$ and $\beta_i^{(A)}$ are the approximate values of the p_i, α_i and β_i respectively, the rough lower and upper bounds of p_i can be calculated by

$$\begin{cases} p_i^{(\text{lower})} = \max(p_i^{(\text{A})} - 0.2,\ 0) \\ p_i^{(\text{upper})} = \min(p_i^{(\text{A})} + 0.2,\ 1), \end{cases} \tag{17}$$

where $p_i^{(\text{lower})}$ and $p_i^{(\text{upper})}$ are the rough lower and upper bounds of p_i respectively.

The rough lower bound of α_i can be calculated by

$$\alpha_i^{(\text{lower})} = \begin{cases} \dfrac{\alpha_{i-1}^{(A)} + \alpha_i^{(A)}}{2}, & i > 1 \\ 0.9\alpha_i^{(A)}, & i = 1, \end{cases} \tag{18}$$

where $\alpha_i^{(\text{lower})}$ is the rough lower bound of α_i. The rough upper bound of α_i can be calculated by

$$\alpha_i^{(\text{upper})} = \begin{cases} \dfrac{\alpha_{i+1}^{(A)} + \alpha_i^{(A)}}{2}, & i < n \\ 1.1\alpha_i^{(A)}, & i = n, \end{cases} \tag{19}$$

where $\alpha_i^{(\text{upper})}$ is the rough upper bound of α_i.

The rough lower and upper bounds of β_i can be calculated by

$$\begin{cases} \beta_i^{(\text{lower})} = \max(0.5\beta_i^{(A)},\ 0) \\ \beta_i^{(\text{upper})} = 1.5\beta_i^{(A)}, \end{cases} \tag{20}$$

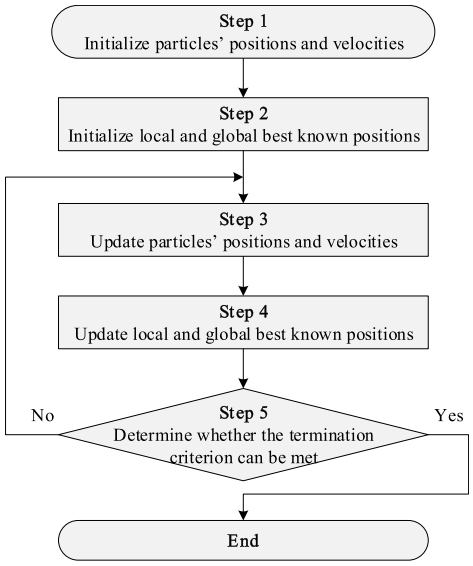


FIGURE 1. PSO flowchart of the LSE based parameter estimation.

where $\beta_i^{(\text{lower})}$ and $\beta_i^{(\text{upper})}$ are the rough lower and upper bounds β_i respectively.

B. STEP BY STEP PROCEDURE

In this study, particle positions are used to denote the specific values of the unknown parameter matrices, and particles' velocities are used to denote the change rates of the parameters. We assume that there are totally q particles in the swarm. The position of the lth particle in the$k$th iteration is expressed as

$$\boldsymbol{\theta}^{(l,k)} = \begin{bmatrix} p_1^{(l,k)} & p_2^{(l,k)} & \cdots & p_n^{(l,k)} \\ \beta_1^{(l,k)} & \beta_2^{(l,k)} & \cdots & \beta_n^{(l,k)} \\ \alpha_1^{(l,k)} & \alpha_2^{(l,k)} & \cdots & \alpha_n^{(l,k)} \end{bmatrix}. \quad (21)$$

The velocity of the lth particle in the kth iteration is expressed as

$$\boldsymbol{V}_{\boldsymbol{\theta}}^{(l,k)} = \begin{bmatrix} v_{p_1}^{(l,k)} & v_{p_2}^{(l,k)} & \cdots & v_{p_n}^{(l,k)} \\ v_{\beta_1}^{(l,k)} & v_{\beta_2}^{(l,k)} & \cdots & v_{\beta_n}^{(l,k)} \\ v_{\alpha_1}^{(l,k)} & v_{\alpha_2}^{(l,k)} & \cdots & v_{\alpha_n}^{(l,k)} \end{bmatrix}. \quad (22)$$

The flowchart of the step by step procedure of the PSO based nonlinear LSE based parameter estimation for the mixed Weibull distribution is given in Fig. 1. And the procedure is composed of 5 steps.

1) Initialize particles' positions and velocities;
2) Initialize local and global best known positions;
3) Update particles' positions and velocities;
4) Update the local and global best known positions;
5) Determine whether the termination criterion is met. If it is met, the procedure will be finished; otherwise, go back to update particles' positions and velocities.

1) STEP 1: INITIALIZE PARTICLES' POSITIONS AND VELOCITIES

We use (23) to express the initial values of the weighting factors p_i

$$p_i^{(l,0)} = \frac{p_i^{(\text{lower})} + (p_i^{(\text{upper})} - p_i^{(\text{lower})}) \times rand}{\sum_{j=1}^{n} p_j^{(\text{lower})} + (p_j^{(\text{upper})} - p_j^{(\text{lower})}) \times rand}, \quad (23)$$

where $rand$ is random number generated from the uniform distribution $U(0, 1)$. Equation (23) can make sure that the summation of all $p_i^{(l,0)}$ $(i = 1, 2, \cdots, n)$ equal 1.

The initial values of the shape parameters β_i can be expressed as

$$\beta_i^{(l,0)} = \beta_i^{(\text{lower})} + (\beta_i^{(\text{upper})} - \beta_i^{(\text{lower})}) \times rand. \quad (24)$$

And the initial values of the scale parameters α_i can be expressed as

$$\alpha_i^{(l,0)} = \alpha_i^{(\text{lower})} + (\alpha_i^{(\text{upper})} - \alpha_i^{(\text{lower})}) \times rand. \quad (25)$$

To the initial velocities, we have

$$\begin{cases} v_{p_i}^{(l,0)} = (p_i^{(\text{lower})} - p_i^{(\text{upper})}) + 2(p_i^{(\text{upper})} - p_i^{(\text{lower})}) \times rand \\ v_{\beta_i}^{(l,0)} = (\beta_i^{(\text{lower})} - \beta_i^{(\text{upper})}) + 2(\beta_i^{(\text{upper})} - \beta_i^{(\text{lower})}) \times rand \\ v_{\alpha_i}^{(l,k)} = (\alpha_i^{(\text{lower})} - \alpha_i^{(\text{upper})}) + 2(\alpha_i^{(\text{upper})} - \alpha_i^{(\text{lower})}) \times rand. \end{cases} \quad (26)$$

We can obtain the initial particles' positions $\boldsymbol{\theta}^{(l,0)}$ via (23) (24) and (25), as well as the initial particles' velocities $\boldsymbol{V}_{\boldsymbol{\theta}}^{(l,0)}$ via (26).

2) STEP 2: INITIALIZE LOCAL AND GLOBAL BEST KNOWN POSITIONS

We use $\boldsymbol{pbest}^{(l,k)}$ to express the local best known position of the lth particle until the kth iteration, and $\boldsymbol{gbest}^{(k)}$ to express the global best known position up to the kth iteration. As the initial local best known positions are just the initial positions, we have

$$\boldsymbol{pbest}^{(l,0)} = \boldsymbol{\theta}^{(l,0)}, \quad l = 1, 2, \cdots, q. \quad (27)$$

And, we can obtain the initial global best known position $\boldsymbol{gbest}^{(0)}$ by (9) which is also the fitness function in the PSO algorithm.

3) STEP 3: UPDATE PARTICLES' POSITIONS AND VELOCITIES

The updating of the velocity for the lth particle in the$k$th iteration is expressed as

$$\boldsymbol{V}_{\boldsymbol{\theta}}^{(l,k)} = w^{(k)} \times \boldsymbol{V}_{\boldsymbol{\theta}}^{(l,k-1)} + c_1 \times rand \times (\boldsymbol{pbest}^{(l,k-1)} - \boldsymbol{\theta}^{(l,k-1)}) + c_2 \times rand \times (\boldsymbol{gbest}^{(k-1)} - \boldsymbol{\theta}^{(l,k-1)}). \quad (28)$$

where c_1 and c_2 are constants and their values are usually equal to 2, $w^{(k)}$ is the inertia weight of the velocity in the kth iteration. Eberhart and Shi [30], [31] have illustrated that

an improved performance can be acquired when the inertia weight typically decrease linearly from 0.9 to 0.4 during one evolution. We use the following expression to calculate $w^{(k)}$

$$w^{(k)} = 0.4 + 0.5 \times (1 - 10^{-a})^k. \tag{29}$$

where a is a positive integer. As the number of iterations is not set as the terminating condition in Step 5, we cannot express the inertia weight in an absolutely linear way. However, equation (29) is approximately linear when k is less than 1000 and a is set as 3 in the case study given in Section IV.

And the updating of the position for the lth particle in the$k$th iteration is expressed as

$$\boldsymbol{\theta}^{(l,k)} = \boldsymbol{\theta}^{(l,k-1)} + \boldsymbol{V}_{\boldsymbol{\theta}}^{(l,k)}. \tag{30}$$

The parameters' values may violate the constraints given in (6) after updating particles' positions. In this case, we will generate a feasible position randomly according to the initialization method of positions given in Step 1.

4) STEP 4: UPDATE LOCAL AND GLOBAL BEST KNOWN POSITIONS

The local and global best known positions are updated according to the fitness function of the present positions of the particles. In the kth iteration, the updating of the local best known position for the lth particle can be expressed as

$$\boldsymbol{pbest}^{(l,k)} = \begin{cases} \boldsymbol{\theta}^{(l,k)}, & f(\boldsymbol{\theta}^{(l,k)}) < f(\boldsymbol{pbest}^{(l,k-1)}) \\ \boldsymbol{pbest}^{(l,k-1)}, & otherwise. \end{cases} \tag{31}$$

To update the global best known position, we have to do calculations for all ($l = 1$ to q) particles according to the following expression.

$$\boldsymbol{gbest}^{(k)} = \begin{cases} \boldsymbol{pbest}^{(l,k)}, & f(\boldsymbol{pbest}^{(l,k)}) < f(\boldsymbol{gbest}^{(k-1)}) \\ \boldsymbol{gbest}^{(k-1)}, & otherwise. \end{cases} \tag{32}$$

After the q comparisons have been finished, we can get the updated global best known position.

5) STEP 5: DETERMINE WHETHER THE TERMINATION CRITERION CAN BE MET

The procedure will terminate as the fitness function converges and approaches an extremely small value. When the fitness function converges, the difference of fitness function value between the two neighbored iterations will be negligible. Namely, $f(\boldsymbol{gbest}^{(k)})$and $f(\boldsymbol{gbest}^{(k-1)})$ are extremely close.

Therefore, the termination criterion can be expressed as

$$\begin{cases} f(\boldsymbol{gbest}^{(k)}) < \varepsilon_1 \\ |f(\boldsymbol{gbest}^{(k)}) - f(\boldsymbol{gbest}^{(k-1)})| < \varepsilon_2, \end{cases} \tag{33}$$

where ε_1 and ε_2 are both arbitrarily small positive reals.

If (33) can be satisfied, the procedure will terminate, and the elements of $\boldsymbol{gbest}^{(k)}$ will be the ultimate parameter estimates. Otherwise, the procedure will go back to Step 3 and a new iteration will start.

One of PSO's the deficiencies is that its result might be trapped into a local optimal solution. In this case, the fitness function will converge but it cannot approach the extremely small value. To avoid being trapped into the local optimal solution, a predetermined number of iterations is given, and the evolution of the procedure will terminate when the iterations reach the predetermined number. And the results obtained from this evolution will be discarded.

IV. CASE STUDY AND DISCUSSIONS

A. CASE STUDY

To illustrate the application of our method, 200 random numbers generated from the mixed Weibull distribution whose distribution function is given in (34) are used as lifetime samples.

$$F(t) = 1 - 0.2e^{-\left(\frac{t}{800}\right)^{0.5}} - 0.6e^{-\left(\frac{t}{1200}\right)^{3}} - 0.2e^{-\left(\frac{t}{2000}\right)}. \tag{34}$$

To get the 200 random numbers, 40 samples are generated from the Weibull distribution with $\alpha = 800$ and $\beta = 0.5$, 120 samples are generated from the Weibull distribution with $\alpha = 1200$ and $\beta = 3$, and 40 are generated from the Weibull distribution with $\alpha = 2000$ and $\alpha = 1$. To each single Weibull distribution, its random numbers are generate via the Inverse transform technique [21]. The lifetime samples are given in Table 1. The totally 200 samples are divided into 10 groups.

By transforming each lifetime sample and the corresponding empirical distribution function via (12), we can find three straight lines given in Fig. 2.

And we can get the approximate value of each parameter by the methods proposed in Section III. Additionally, the upper and lower limit of each parameter can be determined according to the approximate value. The approximate values and the corresponding upper and lower limits are given in Table 2.

In the case study, we let $\varepsilon_1 = 0.01$, $\varepsilon_2 = 10^{-10}$ and $w^{(k)} = 0.4 + 0.5 \times 0.999^k$, namely a is set as 3 in (29). The population size is set as 200, and the predetermined number of iterations is set as 1000. We have conducted the PSO based procedure for 20 times, one of the evolutions of the PSO based procedure is shown in Fig. 3, and the iteration number of this evolution is between 500 and 600.

The results of the 20 evolutions of the PSO based procedure are given in Table 3. And the total elapsed time of the 20 evolutions running in our MATLAB program is about 254 seconds. So the elapsed time of one evolution is nearly 12.7 seconds. The computer used was Intel (R) Core (TM) i5-6500 four Core Processor 3.6GHz, 8GB of RAM.

We can see that the number of iterations have reached 1000 but the value of the fitness function has not approached ε_1 in the 14^{th}, 17^{th}, and 20^{th} evolution. The values of these evolution will be discarded, as their results have been trapped in the local optimum solution.

Therefore, we take the average values of the left 17 evolutions as the estimates of the parameters for the mixed Weibull distribution, which are given in Table 4.

The results obtained from the GA based nonlinear LSE method and the PSO based MLE method as well as the corresponding elapsed time are also listed in Table 4.

TABLE 1. The lifetime samples.

Group 1	Group 2	Group 3	Group 4	Group 5	Group 6	Group 7	Group 8	Group 9	Group 10
0.04	156.93	455.19	656.46	771.29	931.55	1071.69	1210.35	1464.79	1947.06
0.16	213.58	461.67	660.15	780.87	936.48	1073.11	1212.89	1480.51	2011.31
0.65	238.93	477.88	662.13	796.42	941.37	1078.95	1216.28	1522.10	2029.42
0.71	241.27	485.77	679.62	800.88	960.34	1079.95	1218.28	1538.80	2036.03
1.08	242.02	497.48	686.04	812.57	963.27	1082.43	1250.51	1562.33	2049.29
4.23	256.27	499.56	687.70	814.22	967.64	1101.69	1260.37	1574.39	2136.87
13.23	267.62	501.69	695.14	819.61	972.73	1102.33	1269.48	1589.60	2339.75
13.68	276.96	508.20	695.39	830.00	992.61	1105.41	1276.94	1611.27	2603.21
16.65	336.51	536.39	704.57	841.11	995.67	1109.73	1277.22	1613.93	2653.15
37.33	337.74	536.54	713.11	846.68	1003.48	1116.43	1280.02	1623.78	3113.77
38.27	340.12	570.09	728.09	852.35	1003.50	1131.32	1301.10	1636.95	3419.07
39.95	372.67	570.68	735.54	860.40	1007.51	1132.75	1308.48	1645.97	3546.67
44.14	378.09	572.68	738.21	868.16	1010.89	1139.91	1327.76	1674.33	3902.81
53.14	388.69	579.68	742.56	872.17	1013.30	1157.65	1333.51	1676.37	4154.03
62.28	402.17	590.82	751.60	875.44	1025.50	1167.49	1336.84	1695.49	5034.93
81.43	407.45	606.26	752.76	878.94	1039.43	1177.10	1339.98	1728.19	5148.01
90.94	413.96	634.10	753.24	885.21	1040.68	1198.09	1381.50	1745.97	5556.95
107.67	419.51	637.22	757.59	899.22	1044.07	1203.27	1389.92	1753.99	6202.39
122.76	428.06	639.88	759.81	909.98	1056.60	1207.56	1403.89	1789.56	9541.45
127.59	443.82	645.68	771.26	926.22	1071.33	1207.76	1450.88	1865.91	19024.69

TABLE 2. Parameters' approximate values and bounds.

Values or bounds	p_1	p_2	p_3	α_1	α_2	α_3	β_1	β_2	β_3
Approximate values	0.18	0.72	0.1	214	1102	3546	0.37	1.75	0.5
Upper bounds	0.38	0.92	0.3	658	2324	3901	0.56	2.63	0.75
Lower bounds	0	0.52	0	193	658	2324	0.19	0.88	0.25

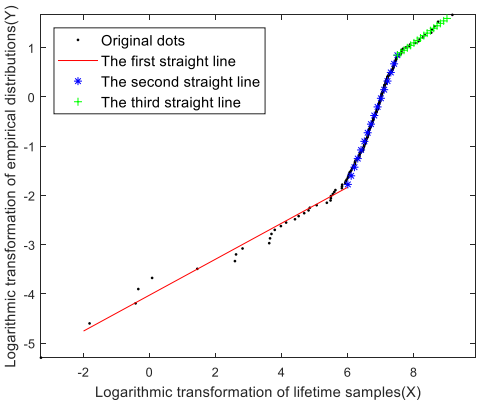


FIGURE 2. Plotting of lifetime samples.

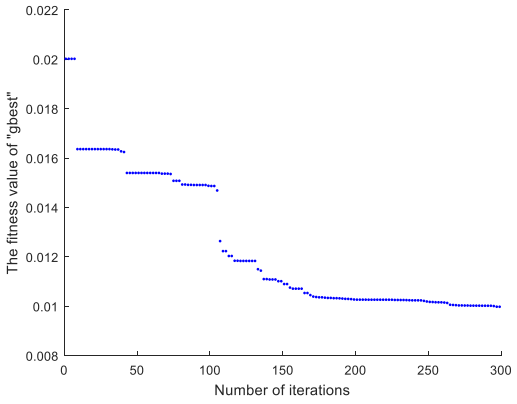


FIGURE 3. One evolutionary process of the PSO based procedure.

In the GA based nonlinear LSE method, the population size is set as 200 which is identical to the PSO based method. The crossover rate is 0.6, and the mutation rate is 0.01. The optimum individual of each generation has been saved and not be involved in the selection, crossover and mutation operation of the next generation.

The CDF obtained by these methods are shown in Fig. 4 to compare with the empirical distribution function. Fig. 4. Shows the CDF obtained by these methods as well as the empirical distribution function.

Fig. 5 Shows curves of the errors between CDF obtained by different methods and the empirical distribution function.

TABLE 3. The results of the twenty evolutions.

NO.	p_1	p_2	p_3	α_1	α_2	α_3	β_1	β_2	β_3	Iterations	Fitness
1	0.255	0.486	0.260	730.342	1027.122	1464.527	0.451	3.026	2.063	319	0.0097
2	0.202	0.455	0.343	772.142	1048.228	1294.589	0.359	3.066	1.692	500	0.0098
3	0.286	0.639	0.075	952.059	1040.546	1853.309	0.470	2.850	1.263	178	0.0099
4	0.223	0.654	0.123	624.655	1062.499	1967.990	0.407	2.693	1.898	317	0.0099
5	0.267	0.463	0.270	871.224	1013.669	1386.853	0.456	2.936	2.334	249	0.0097
6	0.262	0.610	0.128	776.225	1036.601	1802.600	0.452	2.815	2.806	299	0.0097
7	0.174	0.602	0.224	421.988	1064.285	1635.823	0.348	2.789	1.357	342	0.0100
8	0.225	0.667	0.108	602.324	1066.817	2185.930	0.425	2.685	1.949	233	0.0100
9	0.212	0.413	0.375	805.107	1042.700	1262.339	0.363	3.043	1.855	244	0.0099
10	0.228	0.664	0.108	625.667	1058.867	2134.334	0.422	2.696	2.244	263	0.0099
11	0.259	0.678	0.064	829.444	1062.372	2135.438	0.455	2.724	1.301	278	0.0099
12	0.216	0.612	0.172	552.518	1054.958	1813.733	0.417	2.782	1.730	236	0.0099
13	0.254	0.644	0.102	752.117	1053.051	1938.781	0.442	2.767	0.400	243	0.0099
14	0.289	0.711	0.000	1132.105	1105.422	0.000	0.476	2.608	1.958	1001	0.0154
15	0.156	0.557	0.287	433.998	1067.797	1440.651	0.319	2.907	1.290	481	0.0100
16	0.220	0.577	0.203	648.976	1050.355	1578.455	0.402	2.770	1.800	398	0.0097
17	0.162	0.765	0.072	230.506	1095.272	5000.000	0.419	2.484	2.208	1001	0.0124
18	0.256	0.658	0.086	842.391	1050.012	1961.806	0.437	2.715	1.328	249	0.0100
19	0.220	0.662	0.118	570.359	1064.495	2146.839	0.420	2.687	1.893	314	0.0100
20	0.289	0.711	0.000	1132.105	1105.422	5000.000	0.476	2.608	0.000	1001	0.0154

TABLE 4. Parameters' estimated values.

Methods	p_1	p_2	p_3	α_1	α_2	α_3	β_1	β_2	β_3	Elapsed time
Our method	0.23	0.59	0.18	695	1051	1765	0.41	2.82	1.72	12.7 seconds
GA based nonlinear LSE method	0.21	0.60	0.19	658	1124	2415	0.46	2.63	0.75	25 seconds
PSO based MLE method	0.22	0.61	0.17	726	1340	1860	0.6	2.73	1.30	300 seconds

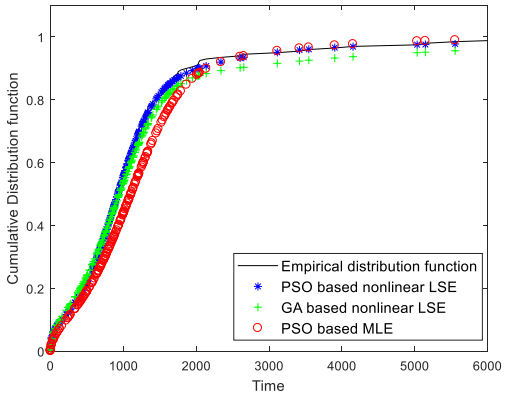


FIGURE 4. The curves of CDF obtained by different methods.

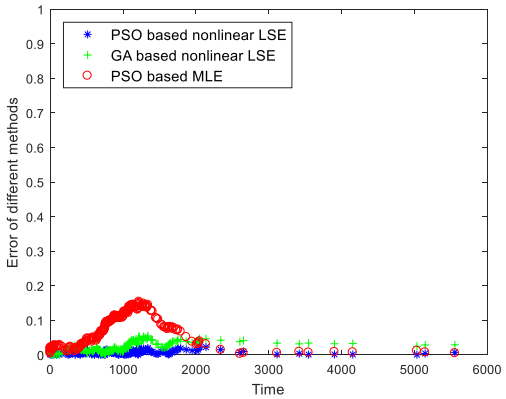


FIGURE 5. The errors between CDF obtained by different methods and the empirical distribution function.

The average errors and maximum errors of the CDF obtained by different methods are listed in Table 5.

B. RESULT DISCUSSIONS

Table 4, Table 5 and Fig. 5 illustrate that both PSO and GA based LSE method can get the satisfied estimation of each parameter, and their results are close to each other. However, our PSO based method is still more accurate and efficient than the GA based nonlinear LSE method. There might be two reasons:

TABLE 5. Errors of different methods.

Methods	The average error	The maximum error
Our method	0.0067	0.0237
GA based nonlinear LSE method	0.0222	0.055
PSO based MLE method	0.0742	0.1538

1) The parameters' estimated values of the GA based nonlinear LSE method are restricted in the initial bounds, since the value of each variable is constrained in a given bound in the GA based optimization. When the true value of the parameter is not located in the given bound, the GA based method cannot find the optimal result. For example, the true values of β_2 and β_3 are both greater than their upper bounds given in Table 2. Their estimated values obtained by the GA based nonlinear LSE method are 2.63 and 0.75 respectively, which are their upper bounds and less their true values.

2) The optimization of the PSO algorithm is realized by the movements of particles, which are only guided by their own best known position as well as the entire swarm's best known position. PSO is simpler than GA as it has no selection, crossover and mutation operations.

Additionally, Table 4 also shows our method is much more efficient than the MLE method, there are also two reasons:

1) The expression of the likelihood function is more complicated than the expression of the summation of the squared residuals. The likelihood function of the mixed Weibull distribution is expressed by

$$L(\alpha_i, \beta_i, p_i) = \prod_{j=1}^{m}\left[\sum_{i=1}^{n} p_i \frac{\beta_i}{\alpha_i^{\beta_i}} t_j^{\beta_i - 1} e^{-\left(\frac{t_j}{\alpha_i}\right)^{\beta_i}}\right]. \quad (35)$$

For single distributions, the likelihood function can be transferred to a simplified formation by taking the natural logarithm. However, the PDF of a mixed distribution is expressed as the weighted summation of several single PDFs, thus it cannot be simplified by taking the logarithm. The natural logarithm of the likelihood function is expressed as

$$L(\alpha_i, \beta_i, p_i) = \sum_{j=1}^{m} \ln\left[\sum_{i=1}^{n} p_i \frac{\beta_i}{\alpha_i^{\beta_i}} t_j^{\beta_i - 1} e^{-\left(\frac{t_j}{\alpha_i}\right)^{\beta_i}}\right]. \quad (36)$$

And the summation of the squared residuals of the mixed Weibull distribution is expressed by

$$f(\alpha_i, \beta_i, p_i) = \sum_{j=1}^{m}\left\{\hat{F}(t_j) - \left[1 - \sum_{i=1}^{n} p_i e^{-\left(\frac{t}{\alpha_i}\right)^{\beta_i}}\right]\right\}^2. \quad (37)$$

Obviously, the calculation of (37) is much simpler than the calculation of both (35) and (36).

2) The likelihood function's order of the magnitude is extremely small, as it is calculated by continually multiplying the probability density functions, particularly when the sample size is large. Therefore, it will cost much time to get the optimal value when the MLE is applied.

V. CONCLUSIONS

The optimization model of the nonlinear LSE based parameter estimation is developed for the mixed Weibull distribution. The objective function of the model is expressed by the summation of the squared nonlinear residuals, which is the difference between the nonlinear CDF and the empirical distribution of the mixed Weibull distribution. In this way, the cumbersome work of transforming the CDF of the mixed Weibull distribution to a linear formation can be avoided.

An approach of determining parameters' approximate values and rough bounds is presented for selecting good starting points used in the PSO. The good starting points can improve the efficiency of PSO algorithm and make it converge faster.

The PSO solution of the LSE is proposed by a step by step procedure. The procedure will terminate as the fitness function converges and approaches an extremely small value. A predetermined number is also treated as the up limit of the procedure's iterations, the evolution of the procedure will terminate when the iterations reach the predetermined number. Hence, the deficiency of being trapped into the local optimal solution can be avoided.

REFERENCES

[1] *Safety Assessment of Transport Airplanes in Commercial Service*, Standard ARP5150A, SAE International Group, Warrendale, PA, USA, 2019.

[2] Z. He, Y. Wang, J. Li, S. Gong, and S. Grzybowski, "New mixed Weibull probability distribution model for reliability evaluation of paper-oil insulation," *Przegląd Elektrotechniczny*, vol. 89, no. 1, pp. 202–206, 2013.

[3] D. Ling, H.-Z. Huang, S. Wang, C.-H. Wu, S.-P. Zhu, and X.-L. Zhang, "A method for multiaxial fatigue reliability analysis based on mixed Weibull distribution," in *Proc. ASME Int. Design Eng. Tech. Conf., Comput. Inf. Eng. Conf.*, Portland, OR, USA, Aug. 2013.

[4] E. Carlucci and L. Tognarelli, "Mixed Weibull distribution as best representative of forced outage distribution to be implemented in blocksim," in *Proc. ASME Power Conf.*, Baltimore, MD, USA, Jul. 2014, Art. no. V001T06A009.

[5] G. P. Du, L. D. He, and B. Zhang, "Reliability modeling of electrical engineering products based on mixed Weibull distribution," *Elect. Meas. Instrum.*, vol. 51, no. 5, pp. 21–25, 2014.

[6] A. Andersen and J. R. Dennison, "Mixed Weibull distribution model of DC dielectric breakdowns with dual defect modes," in *Proc. Annu. Rep. Conf. Elect. Insul. Dielectric Phenomena*, Oct. 2015, pp. 570–573.

[7] R. Ghavijorbozeh and A. Z. Hamadani, "Application of the mixed Weibull distribution in machine reliability analysis for a cell formation problem," *Int. J. Qual. Rel. Manage.*, vol. 34, no. 1, pp. 128–142, 2017.

[8] Z. Yuan, J. Deng, and D. Wang, "Reliability estimation of aero-engine based on mixed Weibull distribution model," in *Proc. 3rd Int. Conf. Adv. Energy Resour. Environ. Eng. (ICAESEE)*, Harbin, China, Dec. 2017, pp. 1–9.

[9] S. Jiang and D. Kececioglu, "Maximum likelihood estimates, from censored data, for mixed-Weibull distributions," *IEEE Trans. Rel.*, vol. 41, no. 2, pp. 248–255, Jun. 1992.

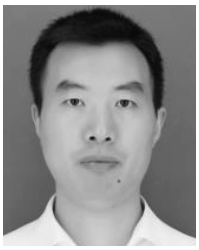
[10] L. Attardi, M. Guida, and G. Pulcini, "A mixed-Weibull regression model for the analysis of automotive warranty data," *Rel. Eng. Syst. Saf.*, vol. 87, no. 2, pp. 265–273, 2005.

[11] Z. Gong, "Estimation of mixed Weibull distribution parameters using the SCEM-UA algorithm: Application and comparison with MLE in automotive reliability analysis," *Rel. Eng. Syst. Saf.*, vol. 91, no. 8, pp. 915–922, 2006.

[12] W.-L. Hung, Y.-C. Chang, and S.-C. Chuang, "Fuzzy classification maximum likelihood algorithms for mixed-Weibull distributions," *Soft Comput.*, vol. 12, no. 10, pp. 1013–1018, Aug. 2008.

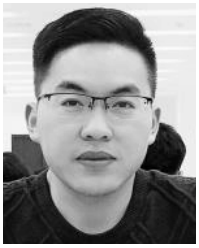
[13] A. E. Touw, "Bayesian estimation of mixed Weibull distributions," *Rel. Eng. Syst. Saf.*, vol. 94, no. 2, pp. 463–473, Feb. 2009.

[14] D. Ling, H.-Z. Huang, and Y. Liu, "A method for parameter estimation of mixed Weibull distribution," in *Proc. IEEE Annu. Rel. Maintainability Symp. (RAMS)*, Fort Worth, TX, USA, Jan. 2009, pp. 129–133.
[15] R. A. Krohling, M. Campos, and P. Borges, "Bare bones particle swarm applied to parameter estimation of mixed Weibull distribution," in *Soft Computing in Industrial Applications: Algorithms, Integration, and Success Stories*, vol. 75. Berlin, Germany: Springer-Verlag, 2010, pp. 53–60.
[16] K. Chi, G. Wang, J. Kang, and K. Wu, "Parameter estimation of mixed Weibull distributions using cuckoo search," *J. Donghua Univ.*, vol. 33, no. 2, pp. 235–238, 2016.
[17] K. Chi, J. Kang, K. Wu, and X. Wang, "Bayesian parameter estimation of Weibull mixtures using cuckoo search," in *Proc. 8th Int. Conf. Intell. Netw. Collaborative Syst. (INCOS)*, Ostrawva, Czech Republic, Sep. 2016, pp. 411–414.
[18] G. Yang, *Life Cycle Reliability Engineering*. Hoboken, NJ, USA: Wiley, 2007.
[19] R. Eberhart and J. Kennedy, "A new optimizer using particle swarm theory," in *Proc. 6th Int. Symp. Micro Mach. Hum. Sci. (MHS)*, Nagoya, Japan, Oct. 1995, pp. 39–43.
[20] V. S. Özsoy, H. H. Örkcü, and H. Bal, "Particle swarm optimization applied to parameter estimation of the four-parameter burr III distribution," *Iranian J. Sci. Technol. Trans. A, Sci.*, vol. 42, no. 2, pp. 895–909, 2018.
[21] H. H. Örkcü, V. S. Özsoy, E. Aksoy, and M. I. Dogan, "Estimating the parameters of 3-p Weibull distribution using particle swarm optimization: A comprehensive experimental comparison," *Appl. Math. Comput.*, vol. 268, pp. 201–226, Oct. 2015.
[22] H. Jiang, J. Wang, J. Wu, and W. Geng, "Comparison of numerical methods and metaheuristic optimization algorithms for estimating parameters for wind energy potential assessment in low wind regions," *Renew. Sustain. Energy Rev.*, vol. 69, pp. 1199–1217, Mar. 2017.
[23] Z. Lu, Y. Sun, and J. Zhou, "Particle swarm optimization based nonlinear least-squares parameter estimation of maintenance time distribution," in *Proc. Int. Colloq. Comput., Commun., Control, Manage. (CCCM)*, Yangzhou, China, 2010, pp. 590–593.
[24] D. Ma, J. Deng, and Z. Zhang, "Comparison and improvements of optimization methods for gas emission source identification," *Atmos. Environ.*, vol. 81, pp. 188–198, Dec. 2013.
[25] D. Ma *et al.*, "Application and improvement of swarm intelligence optimization algorithm in gas emission source identification in atmosphere," *J. Loss Prevention Process Ind.*, vol. 56, pp. 262–271, Nov. 2018.
[26] D. Ma, W. Tan, Z. Zhang, and J. Hu, "Gas emission source term estimation with 1-step nonlinear partial swarm optimization–Tikhonov regularization hybrid method," *Chin. J. Chem. Eng.*, vol. 26, no. 2, pp. 356–363, Feb. 2018.
[27] A. Birolini, *Reliability Engineering: Theory and Practice*, 7th ed. Berlin, Germany: Springer, 2014.
[28] L. F. Zhang, M. Xie, and L. C. Tang, "On weighted least squares estimation for the parameters of Weibull distribution," in *Recent Advances in Reliability and Quality in Design*. London, U.K.: Springer, 2008, pp. 57–84.
[29] Y. Zhang, S. Wang, and G. Ji, "A comprehensive survey on particle swarm optimization algorithm and its applications," *Math. Problems Eng.*, vol. 2015, Feb. 2015, Art. no. 931256.
[30] R. C. Eberhart and Y. Shi, "Comparing inertia weights and constriction factors in particle swarm optimization," in *Proc. Congr. Evol. Comput. (CEC)*, La Jolla, CA, USA, Jul. 2000, pp. 84–88.
[31] Y. Shi and R. C. Eberhart, "Parameter selection in particle swarm optimization," in *Proc. 7th Int. Conf. Evol. Program. (EP)*, San Diego, CA, USA, 1998, pp. 591–600.



ZHONG LU was born in Jiangsu, China, in 1980. He received the bachelor's degree in traffic and transportation engineering and the Ph.D. degree in vehicle operation engineering from the Nanjing University of Aeronautics and Astronautics, Nanjing, China, in 2003 and 2009, respectively.

He has been an Associate Professor with the College of Civil Aviation, Nanjing University of Aeronautics and Astronautics, since 2012. From 2009 to 2012, he was a Lecturer with the College of Civil Aviation, Nanjing University of Aeronautics and Astronautics. His research interests include reliability engineering, system safety assessment, and risk analysis.



LI DONG was born in Henan, China, in 1993. He received the B.S. degree in traffic and transportation engineering from the Nanjing University of Aeronautics and Astronautics, Nanjing, China, in 2017, where he is currently pursuing the master's degree in vehicle operation engineering.

His research interests include lifetime data analysis, reliability modeling, and system safety assessment of civil aircraft.

JIA ZHOU was born in Sichuan, China, in 1981. She received the B.S. degree in applied mathematics from Sichuan Normal University, Chengdu, China, in 2003, and the master's degree in applied mathematics from the Nanjing University of Aeronautics and Astronautics, Nanjing, China, in 2006.

She is currently a Senior Engineer with the Department of Aircraft Maintenance, China Eastern Airlines Jiangsu Ltd. Her research interests include reliability management, failure data analysis, and reliability modeling.