

CASE STUDY

Open Access



# Selecting statistical model and optimum maintenance policy: a case study of hydraulic pump

S. Ruhi<sup>1\*</sup> and M. R. Karim<sup>2</sup>

\*Correspondence:  
sabba.ruhi@gmail.com

<sup>1</sup> Department of Statistics,  
Pabna University of Science  
and Technology, Pabna,  
Bangladesh

Full list of author information  
is available at the end of the  
article

## Abstract

**Introduction:** Proper maintenance policy can play a vital role for effective investigation of product reliability. Every engineered object such as product, plant or infrastructure needs preventive and corrective maintenance.

**Case description:** In this paper we look at a real case study. It deals with the maintenance of hydraulic pumps used in excavators by a mining company. We obtain the data that the owner had collected and carry out an analysis and building models for pump failures. The data consist of both failure and censored lifetimes of the hydraulic pump.

**Discussion and evaluation:** Different competitive mixture models are applied to analyze a set of maintenance data of a hydraulic pump. Various characteristics of the mixture models, such as the cumulative distribution function, reliability function, mean time to failure, etc. are estimated to assess the reliability of the pump. Akaike Information Criterion, adjusted Anderson–Darling test statistic, Kolmogorov–Smirnov test statistic and root mean square error are considered to select the suitable models among a set of competitive models. The maximum likelihood estimation method via the EM algorithm is applied mainly for estimating the parameters of the models and reliability related quantities.

**Conclusions:** In this study, it is found that a threefold mixture model (Weibull–Normal–Exponential) fits well for the hydraulic pump failures data set. This paper also illustrates how a suitable statistical model can be applied to estimate the optimum maintenance period at a minimum cost of a hydraulic pump.

**Keywords:** Case study, EM algorithm, Hydraulic pumps, Maintenance data, Mixture models, Reliability

## Introduction

Every engineered object (product, plant or infrastructure) needs preventive and corrective maintenance. The cost of maintenance can vary from 5 to 30 % (Campbell 1995) of the operating budget depending on the industry sector. This implies that businesses need to manage maintenance effectively to ensure minimum costs. This requires proper data management to assist in building models for effective decision making.

In this paper we look at a real case study. It deals with the maintenance of hydraulic pumps used in excavators by a mining company. We look at the data that the owner

(mining company) had collected and carry out an analysis and build models for pump failures. The data given in Murthy et al. (2015) and Karim et al. (2015) consist of both failure and censored lifetimes of the pump. Murthy et al. (2015) and Karim et al. (2015) showed that the threefold Weibull mixture distribution is the best distribution for the data among the three competing distributions (single Weibull, twofold Weibull mixture and threefold Weibull mixture). In this paper we search a suitable distribution for the data from a set of competitive mixture models (based on Weibull, Exponential, Normal and Lognormal distributions). Finally the selected distribution is used to find out the optimum time at which the expected cost for maintenance of the pump will be minimum.

The remainder of the article is organized as follows: “[Hydraulic pump failure data](#)” section describes a set of hydraulic pump failure data which will be analyzed in this paper. “[Mixture models for modeling failure data](#)” section presents the mixture models for modeling failure data. “[Parameter estimation](#)” section presents the MLEs of the parameters of mixture models by applying the Expectation–Maximization (EM) algorithm. “[Model selection](#)” section describes about the model selection for the data through graphical and statistical approaches. “[Optimum maintenance cost](#)” section expresses a procedure in which we have tried to find out the optimum time at which the expected cost for maintenance of the pump will be minimum. Finally, “[Conclusion](#)” section concludes the article with a discussion of the key findings.

### **Hydraulic pump failure data**

The hydraulic pumps considered here are used in excavators by a mining company. In open cut mines, coal and overburden are transported using excavators and dump trucks. An excavator is a complex machine consisting of several systems. The hydraulic system is one of the important systems comprised of several hydraulic pumps (for linear and rotational motions), hydraulic oil filters and several hydraulic lines. A pump is considered to have failed if it cannot provide the required flow rate at the required pressure. The data recorded by the maintenance department consist of the failure times (for units that have failed and required Corrective Maintenance action) and service times (for units that have not failed yet and were sent for Preventive Maintenance action) for 102 U and presented in Table 1. The column, labeled “Age” means the age (in hours) of the item at the end of the data collection period and the column labeled “Type” indicates whether the data is a failure data (denoted by 1) or censored data (denoted by 0). As can be seen the data consists of 45 failures and 57 censored ages. More detail description of the data can be found in Murthy et al. (2015) and Karim et al. (2015).

### **Mixture models for modeling failure data**

A variety of statistical models have been developed and studied extensively in the analysis of product failure data (Kalbfleisch and Prentice 1980; Meeker and Escobar 1998; Blischke and Murthy 2000; Lawless 2003; Murthy et al. 2004). A set of mixture models that have been used to analyze the pump failure data, given in Table 1, are discussed below.

**Table 1 Hydraulic pump failure data**

Age (h)	Type	Age (h)	Type	Age (h)	Type	Age (h)	Type
81	0	3333	1	9334	1	12,198	0
149	1	3569	1	9368	1	12,198	0
245	1	3837	0	9729	1	12,198	0
340	1	3837	0	9751	0	12,198	0
407	1	4150	0	10,299	1	12,236	0
461	1	5123	1	10,389	0	12,236	0
629	1	5258	1	10,413	0	12,236	0
856	0	5662	0	10,557	1	12,236	0
947	0	5923	1	10,944	1	12,236	0
1460	1	6333	1	10,970	1	12,236	0
1513	1	6717	1	11,647	0	12,394	0
1670	1	7207	1	11,678	1	12,459	0
1688	0	7265	1	11,686	1	13,097	0
2093	0	7624	1	11,798	0	13,497	0
2242	0	7625	0	11,869	0	13,497	0
2242	0	7973	1	11,869	0	13,497	0
2242	0	8183	1	11,923	0	13,497	0
2242	0	8217	1	12,005	0	13,497	0
2242	0	8390	1	12,082	0	13,497	0
2607	1	8462	1	12,090	0	13,497	0
2668	1	8728	1	12,136	0	14,407	1
2806	1	8817	1	12,141	0	15,536	1
3132	0	8870	1	12,143	0	16,289	1
3132	0	8884	0	12,163	0	17,517	1
3132	0	9055	1	12,198	0		
3132	0	9182	1	12,198	0		

The cumulative distribution function (cdf) of a general  $n$ -fold mixture model involves  $n$  subpopulations is given by

$$G(t) = \sum_{i=1}^n p_i F_i(t), \quad t \geq 0 \tag{1}$$

where  $p_i > 0$  and  $\sum_{i=1}^n p_i = 1$ . Here  $F_i(t)$  is the cdf of the  $i$ -th sub-population and  $p_i$  is the mixing probability of the  $i$ -th sub-population. The corresponding probability density function (pdf) is given by

$$g(t) = \sum_{i=1}^n p_i f_i(t), \quad t \geq 0 \tag{2}$$

where  $f_i(t)$  is the pdf associated with  $F_i(t)$ . And the reliability function is

$$R(t) = 1 - G(t) = 1 - \sum_{i=1}^n p_i F_i(t), \quad t \geq 0. \tag{3}$$

The cumulative distribution functions, probability density functions and reliability functions for the various twofold and threefold mixture models can be obtained from Eqs. (1)–(3) by putting  $n = 2$  and  $n = 3$ , respectively. Ruhi et al. (2015) applied a twofold Weibull mixture model for analyzing failure data. More literatures on the applications of mixture models can be found in Titterington et al. (1985), Mendenhall and Hader (1958), Ahmad and Abdelrahman (1994), and Murthy et al. (2004).

**Parameter estimation**

We estimate the parameters of different mixture models by applying the maximum likelihood estimation method. We apply the Expectation–Maximization (EM) algorithm to find the maximum likelihood estimates (MLEs) of the parameters. Details on the application of EM algorithm for mixture models with censored data can be found in Ateya (2012), Bordes and Chauveau (2012) and Ruhi, et al. (2015). Karim, et al. (2015) have applied single Weibull, twofold Weibull mixture and threefold Weibull mixture models for this data set and suggested the threefold Weibull mixture model as the best fitted model on the basis of various graphical and statistical approaches. In addition to threefold Weibull mixture model, here we have assumed two other threefold mixture models (Weibull-Normal-Exponential and Normal-Lognormal-Weibull) for the data. Our aim is to find out whether any other threefold mixture model fits this data set better than the threefold Weibull mixture model or not. And if the distribution changed, what would be its effect on optimal maintenance policy.

The parameters of these three mixture models are estimated by applying maximum likelihood method via the Expectation–Maximization (EM) algorithm. R programming codes are written for all computations of the paper. Programming codes for analyzing the data with Weibull–Normal–Exponential mixture model are given in the “Appendix”. The given codes can be used for other two models after simple modifications, mainly related to the functions `dweibull()`, `pweibull()`, `dnorm()`, `pnorm()`, `dexp()` and `pexp()` and the parameter vector  $\theta$ .

The MLEs of the parameters are displayed in Table 2. In Table 2, the parameters,  $p_1$ ,  $p_2$ , and  $p_3$  represent the mixing probabilities of the 1st, 2nd and 3rd sub-populations, respectively.

**Table 2 MLEs of the parameters of assumed models**

Threefold mixture models	MLEs of parameters
Weibull ( $\beta_1, \eta_1$ )–Weibull ( $\beta_2, \eta_2$ )–Weibull ( $\beta_3, \eta_3$ )	$\{\beta_1, \eta_1, \beta_2, \eta_2, \beta_3, \eta_3, p_1, p_2, p_3\} =$ {1.0191, 2364.0191, 5.5758, 9481.8351, 16.6426, 16535.5039, 0.1659, 0.3220, 0.5120}
Weibull ( $\beta, \eta$ )–Normal ( $\mu, \sigma$ )–Exponential ( $\delta$ )	$\{\beta, \eta, \mu, \sigma, \delta, p_1, p_2, p_3\} =$ {5.5391, 9527.83, 15991.11, 1073.821, 0.0004, 0.3249, 0.5076, 0.1674}
Normal ( $\mu_1, \sigma_1$ )–Lognormal ( $\mu_2, \sigma_2$ )–Weibull ( $\beta, \eta$ )	$\{\mu_1, \sigma_1, \mu_2, \sigma_2, \beta, \eta, p_1, p_2, p_3\} =$ {15992.0308, 1072.7513, 7.5063, 1.3759, 5.4782, 9497.0899, 0.4947, 0.1872, 0.3180}

**Comment**

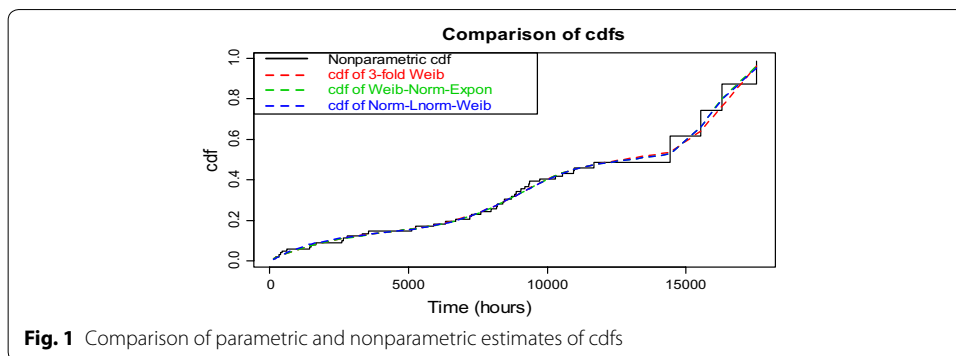
- For Weibull  $(\beta_1, \eta_1)$ –Weibull  $(\beta_2, \eta_2)$ –Weibull  $(\beta_3, \eta_3)$  mixture model, the mean for  $F_3(t; \beta_3, \eta_3) = 16,018.005 >$  mean for  $F_2(t; \beta_2, \eta_2) = 8760.457 >$  mean for  $F_1(t; \beta_1, \eta_1) = 2345.628$ .
- For Weibull  $(\beta, \eta)$ –Normal  $(\mu, \sigma)$ –Exponential  $(\delta)$  mixture model, the mean for  $F_2(t; \mu, \sigma) = 15,991.110 >$  mean for  $F_1(t; \beta, \eta) = 8799.642 >$  mean for  $F_3(t; \delta) = 2500.000$ .
- For Normal  $(\mu_1, \sigma_1)$ –Lognormal  $(\mu_2, \sigma_2)$ –Weibull  $(\beta, \eta)$  mixture model, the mean for  $F_1(t; \mu_1, \sigma_1) = 15,992.031 >$  mean for  $F_3(t; \beta, \eta) = 8765.749 >$  mean for  $F_2(t; \mu_2, \sigma_2) = 4688.418$ .

**Model selection**

This section applies the graphical and statistical approaches for selecting the best fitted model for the data set among three competitive threefold mixture models listed in Table 2. A relatively straightforward approach to select a tentative model is to utilize the plotting methodology where the cdfs obtained from parametric estimates are compared with the empirical distribution function. More detail about this comparison can be found in Blischke et al. (2011). The cdfs of threefold Weibull, Weibull–Normal–Exponential and Normal–Lognormal–Weibull mixture models are compared with the empirical distribution function (nonparametric estimate of cdf from Kaplan–Meier (KM) estimate) and the results are displayed in Fig. 1.

Figure 1 indicates that all the cdfs obtained from the three different mixture models give approximately same result, except at the right tail of the figure of cdfs, where the cdfs of Weibull–Normal–Exponential and Normal–Lognormal–Weibull mixture models belong slightly closer to the nonparametric estimate of cdf than that of the cdf of threefold Weibull mixture model. Hence we may consider both the Weibull–Normal–Exponential and Normal–Lognormal–Weibull mixture models for the data set.

The statistical approaches provide a more rigorous method for model selection and validation. Various statistics [such as adjusted Anderson–Darling (AD\*) value, Kolmogorov–Smirnov (KS) test statistic, Akaike Information Criterion (AIC) and root mean square error (RMSE)] are applied for model selection and validation. The estimates of AIC, AD\*, KS test statistic and RMSE for the three competitive models are given in Table 3.



**Fig. 1** Comparison of parametric and nonparametric estimates of cdfs

**Table 3 Estimates of AIC, AD\*, KS test statistic and RMSE for the models**

Threefold mixture models	AIC	AD*	KS test	RMSE
Threefold Weibull	965.5942	0.6272	0.1068	0.0247
Weibull–Normal–Exponential	963.2532	0.5278	0.0876	0.0209
Normal–Lognormal–Weibull	964.6492	0.4781	0.0877	0.0217

From Table 3, we found that the Weibull–Normal–Exponential mixture model contains the smallest values of AIC and RMSE and the Normal–Lognormal–Weibull mixture model contains the smallest value of AD\* test statistic among all of the mixture models. Hence, it can be concluded that, among these mixture models, Weibull–Normal–Exponential mixture model can be selected as the best model for hydraulic pump failure data according to the values of AIC and RMSE.

We have also applied the Kolmogorov–Smirnov (KS) test statistic as a goodness-of-fit test for these threefold mixture models. At the 5 % level of significance, with  $n = 102$ , the critical value of the Kolmogorov–Smirnov one-sample test is  $1.36/\sqrt{102} = 0.135$  (Siegel and Castellan 1988). Since the observed value of the KS test statistic for all the threefold mixture models (given in Table 3) are less than the critical value, we cannot reject the null hypothesis,  $H_0$ , that the observed data are from a population specified by these threefold mixture distribution. But we may consider that among all these three mixture models the Weibull–Normal–Exponential mixture model gives the smallest value for the KS test statistic.

According to Karim et al. (2015), let us introduce the following notations:

- $q$ : Probability that the pump is scrapped and replaced by a new one under service exchange
- $1 - q$ : Probability that the pump is not scrapped and reconditioned under service exchange
- $p$ : Probability that the item used in service exchange is installed correctly
- $1 - p$ : Probability that the item used in service exchange is not installed correctly
- $F_N(t)$ : Failure distribution of new item installed correctly
- $F_R(t)$ : Failure distribution of reconditioned item installed correctly
- $F_I(t)$ : Failure distribution of incorrectly installed item (new or reconditioned)

It is easily seen (using the conditional approach) that the time to failure of an item used in service exchange is given by a distribution function (Karim et al. 2015)

$$G_3(t) = (1 - p)F_I(t) + (1 - q)pF_R(t) + qpF_N(t) \tag{4}$$

Note that the MTTF (mean time to failure) for a new item installed correctly > MTTF for a reconditioned item installed correctly > MTTF for an item (new or reconditioned) installed incorrectly. If we select the Weibull ( $\beta, \eta$ )–Normal ( $\mu, \sigma$ )–Exponential ( $\delta$ ) mixture model as the best model for the data, then according to the Table 4 of Karim et al. (2015), we can write

$$p_3 = (1 - p), \quad p_1 = (1 - q)p \quad \text{and} \quad p_2 = qp \tag{5}$$

$$F_3(t; \delta) = F_I(t), \quad F_1(t; \beta, \eta) = F_R(t) \quad \text{and} \quad F_2(t; \mu, \sigma) = F_N(t) \tag{6}$$

**Table 4 Optimal  $T^*$  and  $J(T^*)$  for different values of  $\xi$**

Model	Optimal values	Additional cost			
		$\xi = 70,000$	$\xi = 90,000$	$\xi = 110,000$	$\xi = 130,000$
Threefold Weibull	$T^*$	14,631	14,484	14,377	14,295
	$J(T^*)$	10.40593	11.43373	12.45314	13.46734
Weibull–Normal–Exponential	$T^*$	14,468	14,361	14,286	14,230
	$J(T^*)$	10.33359	11.33318	12.3265	13.31607
Normal–Lognormal–Weibull	$T^*$	14,476	14,368	14,291	14,234
	$J(T^*)$	10.32712	11.32669	12.31991	13.30929

Using the estimates of  $p_1, p_2$  and  $p_3$  from Table 2 in Eq. (5), we get the estimates of  $p = 0.8326$  and  $q = 0.6096$ .

**Optimum maintenance cost**

Obtaining the solution to the problem involves building a model and deciding on the optimal age for PM action requires an objective function. The objective function is the asymptotic expected cost per unit time. Note that every time instant an exchanged pump is put into operation can be viewed as a renewal point for a renewal process characterizing the replacements of pumps over time. The time between two successive renewal points defines a cycle. The asymptotic expected cost per unit time can be obtained as the ratio of the expected cycle cost (ECC) and the expected cycle length (ECL).

The time to failure for a pump,  $X$ , is a random variable with distribution function  $F(x)$ . A PM action results if  $X \geq T$  in which case the cycle length is  $T$  with probability  $R(T)$ . A CM action results when  $X < T$  and the cycle length is  $X$ . As a result ECL is given by

$$ECL = \int_0^t tf(t) + TR(T) = \int_0^T R(t)dt \tag{7}$$

Let  $C_f$  and  $C_p$  denote the average cost of a CM and a PM replacement respectively. We will discuss the derivation of this cost later in the section. As a result ECC is given by

$$ECC = C_f F(T) + C_p R(T) \tag{8}$$

From (7) and (8) we have the asymptotic average cost per unit time given by

$$J(T; F(.)) = \frac{C_f F(T) + C_p R(T)}{\int_0^T R(t)dt} \tag{9}$$

$T^*$ , the optimal  $T$ , is the value that yields a minimum for  $J(T; F(.))$ .

The optimal  $T$  depends on the average cost of each CM and PM. Like Karim et al. (2015), we use the following additional notations and assumptions.

$C_n$ : Sale price for new pump (\$80,000).

$C_r$ : Cost (charged by the service agent) for reconditioning a pump under CM or PM action (\$60,000).

$\xi$ : Additional cost (due to downtime, loss in revenue, etc.) resulting from CM action. We look at values of  $\xi = \$70,000, \$90,000, \$110,000$  and  $\$130,000$ .

A maintenance action involves replacement by a new item or a reconditioned item with probabilities  $q$  and  $(1 - q)$  respectively. As a result, the average cost of a PM action is  $C_p = qC_n + (1 - q)C_r$  and of a CM action is  $C_f = C_p + \xi$ . The optimal  $T^*$  is obtained using (9) with threefold mixture cdf  $F(t) = G_3(t)$  and the optimal expected cost per unit time is given by  $J(T^*; F(\cdot))$  i.e.,  $J(T^*; G_3(\cdot))$ .

Here we can see that, the optimal  $T^*$  depend on the additional cost  $\xi$ . The optimal  $T^*$  and optimal expected cost per unit time  $J(T^*)$  on various values of  $\xi$  for the three different threefold mixture models has been estimated. These results are given in Table 4, from where it can be seen, for every model, the optimal  $T^*$  decrease and optimal  $J(T^*)$  increasing with  $\xi$  increases, as to be expected.

Table 4 indicates that the threefold Weibull mixture model gives a bit larger optimal maintenance period  $T^*$  than other two models, however the Weibull–Normal–Exponential model shows a reduction in the maintenance cost than the threefold Weibull mixture model for all  $\xi$ .

## Conclusion

Proper data management (data collection and analysis) is very important for effective maintenance of any engineered object. Data is critical for building and selecting suitable statistical models and model provides new insights for improvements to maintenance operations.

This paper has dealt with a real case study to illustrate how statistical models can be selected and applied for estimating optimum maintenance period and cost of a hydraulic pump. It is recommended that the Weibull–Normal–Exponential mixture model can be selected as the best model for hydraulic pump failure data among three competitive models. This model suggests the optimum maintenance period for the pump that reduces the maintenance cost. Annotated R code is provided for analyzing hydraulic pump failure data with Weibull–Normal–Exponential mixture model. The code can be modified easily to apply other threefold mixture models.

### Authors' contributions

The authors with the consultation of each other have carried out this work and drafted the manuscript together. All authors read and approved the final manuscript.

### Author details

<sup>1</sup> Department of Statistics, Pabna University of Science and Technology, Pabna, Bangladesh. <sup>2</sup> Department of Statistics, University of Rajshahi, Rajshahi 6205, Bangladesh.

### Competing interests

The authors declare that they have no competing interests.



## Appendix

R codes for analyzing pump failure data with Weibull–Normal–Exponential mixture model.

### # A. Function for estimating parameters ---

```

library(splines)
library(survival)

# Observed data log-likelihood function ----
# t: lifetime variable, d: failure/censored indicator; theta: parameter vector
loglik.obs <- function(t, d, theta){
  sum(d*log(theta[6]*dweibull(t, shape=theta[1], scale=theta[2], log = FALSE)
  + theta[7]*dnorm(t, mean=theta[3], sd =theta[4], log = FALSE)
  + theta[8]*dexp(t, rate=theta[5], log = FALSE))
  + (1-d)*log(theta[6]*(1-pweibull(t, shape=theta[1], scale=theta[2], lower.tail=TRUE, log.p=FALSE))
  + theta[7]*(1-pnorm(t, mean=theta[3], sd=theta[4], lower.tail = TRUE, log.p = FALSE))
  + theta[8]*(1-pexp(t, rate=theta[5], lower.tail = TRUE, log.p = FALSE)))
}

# em.tiny: a small value - convergence criterion for EM algorithm
WeibNormExpMix <- function(t, d, theta, em.tiny){
  n <- length(t); K <- 3; tiny <- 10^(-8); change.lik <- 0.05; maxi.em.rep <- 500; em.rep <- 1
  fti <- matrix(0, nrow = n, ncol = K); loglik <- array()

  loglik.old <- loglik.obs(t, d, theta)
  # ----- Iteration for EM algorithm
  while(change.lik > em.tiny && em.rep <= maxi.em.rep) {
    # ----- E-step
    tempj1.f <- dweibull(t, shape=theta[1], scale=theta[2], log = FALSE)
    tempj2.f <- dnorm(t, mean=theta[3], sd =theta[4], log = FALSE)
    tempj3.f <- dexp(t, rate=theta[5], log = FALSE)
    tempj1.c <- (1-pweibull(t, shape=theta[1], scale=theta[2], lower.tail = TRUE, log.p = FALSE))
    tempj2.c <- (1-pnorm(t, mean=theta[3], sd=theta[4], lower.tail = TRUE, log.p = FALSE))
    tempj3.c <- (1-pexp(t, rate=theta[5], lower.tail = TRUE, log.p = FALSE))

    p <- c(theta[6], theta[7], theta[8])
    fti[, 1] <- (p[1]*tempj1.f/(p[1]*tempj1.f + p[2]*tempj2.f+ p[3]*tempj3.f))^d*(p[1]*tempj1.c/(p[1]*tempj1.c +
    p[2]*tempj2.c+ p[3]*tempj3.c))^(1-d)
    fti[, 2] <- (p[2]*tempj2.f/(p[1]*tempj1.f + p[2]*tempj2.f+ p[3]*tempj3.f))^d*(p[2]*tempj2.c/(p[1]*tempj1.c +
    p[2]*tempj2.c+ p[3]*tempj3.c))^(1-d)
    fti[, 3] <- (p[3]*tempj3.f/(p[1]*tempj1.f + p[2]*tempj2.f+ p[3]*tempj3.f))^d*(p[3]*tempj3.c/(p[1]*tempj1.c +
    p[2]*tempj2.c+ p[3]*tempj3.c))^(1-d)

    # ----- M-step
    p1 <- sum(fti[, 1])/n; p2 <- sum(fti[, 2])/n; p3 <- sum(fti[, 3])/n; p <- c(p1, p2, p3)

    t1.new <- array(); t2.new <- array(); t3.new <- array()
    d1.new <- array(); d2.new <- array(); d3.new <- array()
    fj1.new <- array(); fj2.new <- array(); fj3.new <- array()

    # ----- To omit observation from LF if its weight is very small, less than tiny
    j <- 0
    for(i in 1:n){
      if(fti[i,1] >= tiny){
        j <- j+1; t1.new[j] <- t[i]; d1.new[j] <- d[i]; fj1.new[j] <- fti[i,1]
      }
    }
  }
}

```



```

Pr[i] <- 1-temp.prod
}
nf<- 0
for(i in 1:n){
  if(d[i]==1){
    nf <- nf + 1
    tt[nf] <- t[i]
    new.Pr[nf] <- Pr[i] # cdf corresponding to failed observations
  }
}
max.t <- max(t)
for(i in 1:n){
  if((t[i] == max.t) && (d[i]==1)){
    new.Pr[nf] <- new.Pr[nf-1]+(1-new.Pr[nf-1])*0.90 # if largest observation is failure, re-estimate
  }
}
Fnz <- new.Pr # Nonparametric cdf (KM estimate)
Fnz[nf+1] <- 1

# ---- Parametric estimates of CDF
beta.hat <- theta[1]; eta.hat <- theta[2]; mu.hat <- theta[3]; sigma.hat <- theta[4]; delta.hat <- theta[5]
p1.hat <- theta[6]; p2.hat <- theta[7]; p3.hat <- theta[8]
mle.F1 <- pweibull(tt, shape = beta.hat, scale = eta.hat, lower.tail = TRUE, log.p = FALSE)
mle.F2 <- pnorm(tt, mean=mu.hat, sd=sigma.hat, lower.tail = TRUE, log.p = FALSE)
mle.F3 <- pexp(tt, rate=delta.hat, lower.tail = TRUE, log.p = FALSE)
mle.F <- p1.hat*mle.F1 + p2.hat*mle.F2 + p3.hat*mle.F3

MLE.CDF <- mle.F
mle.F[nf+1] <- 0.9999999999999999 # a value close to 1
mle.F.lag <- array() # z[i-1]
mle.F.lag[1] <- 0
for(i in 1:nf){
  mle.F.lag[i+1] <- mle.F[i]
}

Fnz.lag <- array() # Fnz[i-1]
Fnz.lag[1] <- 0
for(i in 1:nf){
  Fnz.lag[i+1] <- Fnz[i]
}
A <- array(); B <- array(); C <- array()
A <- -mle.F- log(1-mle.F) + mle.F.lag + log(1-mle.F.lag)
B <- 2*log(1-mle.F)*Fnz.lag - 2*log(1-mle.F.lag)*Fnz.lag
C <- log(mle.F)*Fnz.lag^2 - log(1-mle.F)*Fnz.lag^2 - log(mle.F.lag)*Fnz.lag^2 + log(1-mle.F.lag)*Fnz.lag^2
C[1] <- 0 # Replace the missing value in the first row of 'C_i' with a zero.
AD.adj <- nf*sum(A,B,C)
return(list("f.time"=c(tt), "KM.F"=c(new.Pr), "MLE.F"=c(MLE.CDF), "AD.adj.value"=AD.adj))
}

theta.hat <- WeibNormExpMix(Age, Type, theta.ini, 10^(-6))$theta.hat # MLE of parameters
AD.adj.KM(Age, Type, theta.hat) # Estimation of AD value

# D. Function for the estimation of AIC, KS test statistic and RMSE ----
AIC.KS.RMSE <- function(t, d, theta.hat){
theta.ini <- c(1.522, 16500, 12500.5, 6500, 0.0009, 0.35, 0.25, 0.40)
loglik <- WeibNormExpMix(t, d, theta.ini, 10^(-6))$likelihood
n.para <- 7 # change No. of independent parameters

```

```

AIC <- -2*loglik + 2*n.para      # AIC value
KS.ts <- max(abs(AD.adj.KM(t, d, theta.hat)$KM.F - AD.adj.KM(t, d, theta.hat)$MLE.F)) #
Kolmogorov-Smirnov test statistic
RMSE <- sd(AD.adj.KM(t, d, theta.hat)$KM.F - AD.adj.KM(t, d, theta.hat)$MLE.F)
      # Root Mean Squared Error
return(list("AIC.hat"=AIC, "KS.hat"=KS.ts, "RMSE.hat"=RMSE))
}

AIC.KS.RMSE(Age, Type, theta.hat)      # Execution of the function

# E. Codes for creating figure ----

plot(AD.adj.KM(Age, Type, theta.hat)$f.time, AD.adj.KM(Age, Type, theta.hat)$KM.F, main="EDF and
Weibull-Normal-Exponential mixture CDF", xlab=" Time (hours)", ylab="CDF", col=1, lty= 1, type="s",
lwd=1)
lines(AD.adj.KM(Age, Type, theta.hat)$f.time, AD.adj.KM(Age, Type, theta.hat)$MLE.F, col=2, lty=2,
type="l", lwd=2)
legend("topleft", c("KM CDF", "Weib-Norm-Expn mixture CDF"), col=c(1,2), text.col=c(1,2), lty=c(1,2),
lwd=2)

# F. Codes for estimating optimal maintenance age by minimizing J(T)

q <- 0.6096 # value of q in the formula
Cn <- 80000; Cr <- 60000; Cp <- q*Cn+(1-q)*Cr
opt.T.JT <- function(zeta){
Cf <- Cp + zeta
TT = seq(4000, 20000, 1) # Possible values of T, opt T would be within TT
JT <- function(T.value){ # ---- J(T)star function
WNE.F1 <- pweibull(T.value, shape=theta.hat[1], scale=theta.hat[2], lower.tail = TRUE, log.p = FALSE)
WNE.F2 <- pnorm(T.value, mean=theta.hat[3], sd=theta.hat[4], lower.tail = TRUE, log.p = FALSE)
WNE.F3 <- pexp(T.value, rate=theta.hat[5], lower.tail = TRUE, log.p = FALSE)
WNE.f <- theta.hat[6]*WNE.F1 + theta.hat[7]*WNE.F2 + theta.hat[8]*WNE.F3
WNE.R <- 1-WNE.F
  fun.t <- function(t){
WNE.f1 <- dweibull(t, shape=theta.hat[1], scale=theta.hat[2], log = FALSE)
WNE.f2 <- dnorm(t, mean=theta.hat[3], sd=theta.hat[4], log = FALSE)
WNE.f3 <- dexp(t, rate=theta.hat[5], log = FALSE)
WNE.f <- theta.hat[6]*WNE.f1 + theta.hat[7]*WNE.f2 + theta.hat[8]*WNE.f3
return(t*WNE.f)
}

int.part <- integrate(fun.t, lower=0, upper = T.value)$value
JT.value <- (Cf*WNE.F + Cp*WNE.R)/(int.part + T.value*WNE.R)
return(JT.value)
}
TT.no <- length(TT); JT.out <- array()
for(j in 1:TT.no){
  JT.out[j] <- JT(TT[j])
}
JT.star.opt <- min(JT.out)
for(i in 1:TT.no){
  if(JT.out[i] == JT.star.opt) {T.est <- TT[i]}
}
return(list(zeta.value=zeta, opt.T=T.est, opt.JT=JT.star.opt))
}

opt.T.JT(130000)

```

Received: 1 December 2015 Accepted: 17 June 2016  
Published online: 04 July 2016

## References

- Ahmad KE, Abdelrahman AM (1994) Updating a nonlinear discriminant function estimated from a mixture of 2-Weibull distributions. *Math Comput Model* 18:41–51
- Ateya SF (2012) Maximum likelihood estimation under a finite mixture of generalized exponential distributions based on censored data. In: *Statistical papers* (5 November 2012), pp 1–15
- Blichke WR, Murthy DNP (2000) *Reliability*. Wiley, New York
- Blichke WR, Karim MR, Murthy DNP (2011) *Warranty data collection and analysis*. Springer, Berlin
- Bordes L, Chauveau D (2012) EM and stochastic EM algorithms for reliability mixture models under random censoring. hal-00685823, v1. <https://hal.archives-ouvertes.fr/hal-00685823v1>
- Campbell JD (1995) Outsourcing in maintenance management—a valid alternative to self-provision. *J Qual Maint Eng* 1(3):18–24
- Kalbfleisch JD, Prentice RL (1980) *The Statistical analysis of failure time data*. Wiley, New York
- Karim MR, Ahmadi A, Murthy DNP (2015) Modeling of maintenance data. In: Presented at ICRESH-ARMS conference, 2015, Lulea, Sweden
- Lawless JF (2003) *Statistical methods for lifetime data*. Wiley, New York
- Meeker WQ, Escobar LA (1998) *Statistical methods for reliability data*. Wiley, New York
- Mendenhall W, Hader RJ (1958) Estimation of parameters of mixed exponentially distributed failure time distributions from censored life test data. *Biometrika* 45:504–520
- Murthy DNP, Xie M, Jiang R (2004) *Weibull models*. Wiley, New York
- Murthy DNP, Karim MR, Ahmadi A (2015) Data management in maintenance outsourcing. *Reliab Eng Syst Saf* 142:100–110
- Ruhi S, Sarker S, Karim MR (2015) Mixture models for analyzing product reliability data: a case study. *SpringerPlus* 4:634
- Siegel S, Castellan NJ (1988) *Nonparametric statistics for the behavioral sciences*. McGraw-Hill, New York
- Titterton M, Smith AFM, Makov UE (1985) *Statistical analysis of finite mixture distribution*. Wiley, New York

**Submit your manuscript to a SpringerOpen<sup>®</sup> journal and benefit from:**

- Convenient online submission
- Rigorous peer review
- Immediate publication on acceptance
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

---

Submit your next manuscript at ► [springeropen.com](http://springeropen.com)

---