Comparison Study on General Methods for Modeling Lifetime Data with Covariates

Haitao Liao and Samira Karimi
Department of Industrial Engineering
University of Arkansas
Fayetteville, Arkansas, USA
liao@uark.edu

***Abstract*—Lifetime data with covariates (e.g., temperature, humidity, and electric current) are frequently seen in engineering applications. An important example is accelerated life testing (ALT) data. In ALT, test units of a product are exposing to severer-than-normal conditions to expedite product failure. The resulting lifetime and/or censoring data with covariates are often modeled by a probability distribution along with a life-stress relationship. However, if the probability distribution and the life-stress relationship selected cannot adequately describe the underlying failure process, the resulting reliability prediction will be misleading. This paper develops a new method for modeling lifetime data with covariates using phase-type (PH) distributions and a general life-stress relationship formulation. A numerical study is presented to compare the performance of this method with a mixture of Weibull distributions model. This general method creates a new avenue to modeling and interpreting lifetime data with covariates for situations where the data-generating mechanisms are unknown or difficult to analyze using existing statistical tools.**

Keywords*—*accelerated life testing; phase-type distributions

I. INTRODUCTION

A covariate (e.g., temperature, humidity, and electric current) is a variable that is possibly predictive of the outcome under study. Data with covariates are frequently seen in engineering application. In particular, accelerated life testing (ALT) data are usually conducted by exposing test units of a product to severer-than-normal conditions to expedite the failure process. The resulting lifetime and/or censoring data are often modeled by a probability distribution along with a life-stress relationship. However, if the probability distribution and life-stress relationship selected cannot adequately describe the underlying failure process, the resulting reliability prediction will be misleading.

In practice, it is natural that there are underlying processes going through a series of stages before failures occur, and many of these processes are often partially or completely unobservable due to technology barriers or lack of understanding of failure mechanisms [1]. Despite some model-selection guidelines [2–4], choosing adequate distributions to fit such data is always a challenging task. Although some commercial software packages provide several options for probability distributions or even a distribution-selection wizard based on the likelihood values of a limited number of candidate models, no generic model-construction and -selection methods have been reported in the related literature. To seek new mathematical and statistical tools to facilitate the modeling of such data, a critical question to be asked is: Can we find a family of versatile probability distributions along with a general life-stress relationship to model complex lifetime data with covariates?

A continuous phase-type (PH) distribution describes the time to absorption of a continuous-time Markov chain (CTMC) defined on a finite-state space. Since the class of PH distributions is dense, any distribution defined on $[0, \infty)$, in principle, can be approximated arbitrarily closely by a PH distribution [5, 6]. In reliability engineering, Ruiz-Castro et al. [7] investigated a repairable cold-standby system using a quasi-birth-and-death process. Jonsson et al. [8] used PH distributions to deal with non-exponential lifetime distributions in a system. Recently, to model ALT data, Liao and Guo [9] explored a new accelerated failure time model [10] based on Erlang-Coxian distributions [11] to characterize ALT data. The ALT model belongs to an accelerated failure time (AFT) model, which incorporates the effect of a covariate on the product's reliability through time scaling.

In this paper, a more general method is proposed for modeling lifetime data with covariates. In this method, a general life-stress relationship is introduced into the Coxian distribution and a maximum likelihood-based approach is utilized to estimate the model parameters and perform model selection. A comparison study is conducted to illustrate the modeling capability of this method. The unique contribution of this work is that it provides a flexible method for modeling and interpreting lifetime data with covariates when the underlying failure mechanisms are unknown or difficult to analyze using the traditional statistical tools.

The remainder of this paper is organized as follows. Section II describes the proposed model based on the Coxian distribution and the widely used model based on a mixture of Weibull distributions. The corresponding model selection aspects are provided on Section III. In Section IV, a numerical example is provided to compare the performance of the proposed method and the mixture of Weibull distributions model. Finally, conclusions are drawn in Section V.

This work is sponsored by the U.S. National Science Foundation under grant CMMI-1635379.

II. General Methods for Modeling Lifetime Data with Covariates

A. Use of a Mixture of Weibull Distributions

A mixture of Weibull distributions is widely used in modeling complex lifetime data that may not be well described by a single probability distribution such as the two- or three-parameter Weibull distribution, lognormal distribution, and Gamma distribution. Moreover, it has also been used to approximate other probability distributions.

The probability density function (PDF) $f(t;\underline{\theta})$ of a mixture of m Weibull distributions can be expressed as:

$$f(t;\underline{\theta}) = \sum_{i=1}^{m} p_i \frac{\beta_i}{\eta_i}\left(\frac{t}{\eta_i}\right)^{\beta_i - 1} \exp\left(-\left(\frac{t}{\eta_i}\right)^{\beta_i}\right), \quad (1)$$

where $\sum_{i=1}^{m} p_i = 1$, β_i and η_i are the shape and scale parameters of the individual Weibull distribution, and $\underline{\theta}$ represent a vector containing all the model parameters. The corresponding cumulative distribution function (CDF) $F(t;\underline{\theta})$ is:

$$F(t;\underline{\theta}) = 1 - \sum_{i=1}^{m} p_i \exp\left(-\left(\frac{t}{\eta_i}\right)^{\beta_i}\right). \quad (2)$$

To quantify the effects of a covariate z on the model parameters, a general approach is to model each individual parameter as a function of the covariate as:

$$F(t;\underline{\tilde{\iota}},-, \quad \sum_{i=1}^{m} p_i \exp\left(-\left(\frac{t}{\eta_i(z;\underline{\alpha}_i)}\right)^{\beta_i(z;\underline{\gamma}_i)}\right). \quad (3)$$

For example, a widely used life-stress relationship takes a log-linear form $\eta_i(z;\underline{\alpha}_i) = \exp(\alpha_{0i} + \alpha_{1i} z)$ and assumes a constant shape parameter $\beta_i(z;\underline{\gamma}_i) = \beta_i$.

B. Use of Coxian Distributions

We consider a CTMC $\{X(t)\}_{t\geq 0}$ with finite states {1, 2, …, N, N+1}, where state N+1 is the only absorbing state and the others are transient states. Let T be the time to absorption of the CTMC. The Coxian distribution is a versatile mathematical model that describes the absorbing time of a CTMC. Fig. 1 shows a three-phase Coxian model.

The infinitesimal generator of $\{X(t)\}_{t\geq 0}$ is given by:

$$\boldsymbol{Q} = \begin{bmatrix} \boldsymbol{S} & \boldsymbol{q} \\ \boldsymbol{0} & 0 \end{bmatrix}. \quad (4)$$

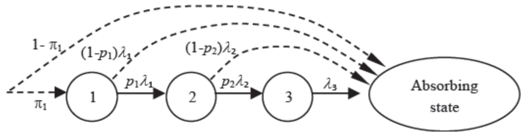

Fig. 1. CTMC for a three-phase Coxian distribution.

For an acyclic PH distribution (e.g., Coxian), $\boldsymbol{S}$ is an upper triangular matrix, $\boldsymbol{q} = -\boldsymbol{S}\boldsymbol{e}$, and $\boldsymbol{e} = [1,1,...,1]'$. For the N-phase Coxian distribution, we have:

$$\boldsymbol{S} = \begin{bmatrix} -\lambda_1 & \lambda_1 p_1 & 0 & 0 \\ 0 & -\lambda_2 & \lambda_2 p_2 & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & & -\lambda_N \end{bmatrix}; \ \boldsymbol{q} = \begin{bmatrix} \lambda_1(1-p_1) \\ \lambda_2(1-p_2) \\ \vdots \\ -\lambda_N \end{bmatrix}. \quad (5)$$

Let the process $\{X(t)\}_{t\geq 0}$ start in the first phase by making $\pi_1 = 1$. Then, the CDF and PDF of the time to absorption T can be expressed as: $F(t) = 1 - \exp(t\boldsymbol{S})\boldsymbol{e}$ and $f(t) = \exp(t\boldsymbol{S})\boldsymbol{q}$, respectively, where $\exp(t\boldsymbol{S})$ represents matrix exponential. A general life-stress relationship can be introduced as follows:

$$\lambda_i(z) = \Psi_i(z;\underline{\Upsilon}_i), \quad (6)$$

where each transition rate λ_i is modeled as a function of the covariate with parameter $\underline{\Upsilon}_i$. Because all λ_i, $i = 1,2,...,N$, are positive, a useful candidate is $\Psi_i(z;\underline{\Upsilon}_i) = \exp(\Upsilon_{0i} + \Upsilon_{1i} z)$. In this paper, this exponential form of life-stress relationship is adopted.

III. Model Selection and Parameter Estimation

A. Maximum Likeliood Estimation Method

To estimate the model parameters $\underline{\tilde{\iota}}$ for a mixture of m Weibull distributions, the maximum likelihood estimation method can be used. Given a set of lifetime data collected under J levels of the covariate $\{t_{jk}, \delta_{jk}, z_j\}_{j=1,2,...,J}$, where $\delta_{jk} = \{1, \text{if } t_{jk} \text{ is lifetime; } 0, \text{otherwise}\}$, the likelihood function can be expressed as:

$$l(\underline{\tilde{\iota}}, \quad \prod_{j=1}^{J}\prod_{k=1}^{K_i}\left[\sum_{i=1}^{m} \begin{array}{l} p_i \dfrac{\beta_i(z_j;\underline{\gamma}_i)}{\eta_i(z_j;\underline{\alpha}_i)}\left(\dfrac{t_{jk}}{\eta_i(z;\underline{\alpha}_i)}\right)^{\beta_i(z_j;\underline{\gamma}_i)-1} \\ \times\exp\left(-\left(\dfrac{t_{jk}}{\eta_i(z_j;\underline{\alpha}_i)}\right)^{\beta_i(z_j;\underline{\gamma}_i)}\right) \end{array}\right]^{\delta_{jk}} \times\left[\sum_{i=1}^{m} p_i \exp\left(-\left(\frac{t_{jk}}{\eta_i(z_j;\underline{\alpha}_i)}\right)^{\beta_i(z_j;\underline{\gamma}_i)}\right)\right]^{1-\delta_{jk}}. \quad (7)$$

where K_j is the total number of observations under the jth level of covariate. Then, the maximum likelihood estimate of $\underline{\theta}$ can be obtained by maximizing the log-likelihood function $\ln l(\underline{\theta})$ after taking the natural-log of (7).

Similarly, to estimate the model parameters in a specific Coxian-based model, the likelihood function is given by:

$$l(\underline{\theta}) = \prod_{j=1}^{J}\prod_{k=1}^{K_j}\left[\exp(t_{jk}\boldsymbol{S}(z_j))\boldsymbol{q}(z_j)\right]^{\delta_{jk}}\left[\exp(t_{jk}\boldsymbol{S}(z_j))\boldsymbol{e}\right]^{1-\delta_{jk}}, \quad (8)$$

where matrices $\boldsymbol{S}(z_j)$ and $\boldsymbol{q}(z_j)$ have the forms given in (5) with all $\lambda_i(z)$, $i=1,2,...,N$, as described by (6). In this research, an expectation-maximization algorithm is utilized to maximize the corresponding log-likelihood function for obtaining the maximum likelihood estimates of the model parameters.

B. Likelihood-based Model Selection

It is worth pointing out that when either the Coxian-based model or the mixture of Weibull distributions is used to model lifetime data with covariates, the model structures are not necessarily pre-determined. Instead, a model selection method can be utilized to determine the models that provide the best fit to the data. In particular, for the Coxian-based model the number of phases can be determined according to the change in the value of maximum likelihood as the number of phases is increased. Similarly, for the mixture of Weibull distributions model the number of Weibull distributions in the model can be obtained by investigating the values of maximum likelihood as more complex models are considered.

An alternative for such model selection problems under the maximum likelihood estimation framework is the use of Akaike Information Criterion (AIC):

$$AIC = 2p - 2\ln l(\hat{\underline{\theta}}), \quad (9)$$

where p is the number of parameters in parameter vector $\underline{\theta}$ of the corresponding model and $\hat{\underline{\theta}}$ is the maximum likelihood estimate of $\underline{\theta}$. The common practice is to choose the model that has the lowest AIC value compared to other candidate models.

C. Construction of Confidence Interval

To quantify the uncertainty in model estimates, a nonparametric bootstrap method can be implemented to construct the interested confidence intervals (CI). This sample-reuse method is quite useful when other alternatives, such as parametric bootstrap and Fisher-Information-based methods, are either intractable or insufficiently accurate. Indeed, this is the case for the Coxian-based model presented in this paper.

In this paper, the nonparametric bootstrap procedure developed by Liao et al. [12] is implemented to construct the pointwise confidence intervals for the Cdfs estimated under different levels of covariate(s). The detailed steps can be described as follows:

1. Each sample (say bootstrap sample i) consisting of n_j data points for each level z_j is obtained by sampling, with replacement, from the original lifetime data with covariates.
2. Parameters $\hat{\underline{\theta}}_i$ are estimated based on each bootstrap sample.
3. Compute the corresponding Cdf at time t using the bootstrap estimate $\hat{\underline{\theta}}_i$.
4. Repeat Steps 1-3 for B (say 5000) times to obtain a set of Cdf estimates at time t as:
$$\{\hat{F}_1(t), \hat{F}_2(t), ..., \hat{F}_B(t)\}.$$
5. Sort the set in an ascending order for each desired time t:
$$\{\hat{F}_{[1]}(t), \hat{F}_{[2]}(t), ..., \hat{F}_{[B]}(t)\}.$$
6. Determine the lower and upper bounds of pointwise $100(1-\alpha)$% confidence interval for the Cdf as:
$$[\hat{F}_{[v]}(t), \hat{F}_{[u]}(t)],$$

where $v = [\alpha B/2]^+$ and $u = [(1-\alpha/2)B]^+$.

IV. Numerical Example

In this section, we use the ALT data reported in [13] to illustrate the use of the proposed method for modeling lifetime data with a single covariate.

A. Experimental Setup

The purpose of this ALT experiment is to estimate the reliability of a type of miniature lamps under the use condition: 2 volts. The highest operating voltage of the lamp is 5 volts. Three constant voltage levels were used in the experiment: 5 volts, 3.5 volts, and 2 volts. After standardization ((volts-2)/3), the three voltage levels are: $z_1 = 1$, $z_2 = 0.5$, and $z_3 = 0$. The observed lifetimes and censoring times are presented in Table 1. To avoid making an assumption on the underlying distribution, the proposed Coxian-based model as well as the mixture of Weibull distributions model are utilized to predict the reliability of this type of miniature lamps.

B. Results of Mixture of Weibull Distributions

Figs. 2 and 3 show the estimated CDFs using the Kaplan-Meier estimator, a mixture model of three Weibull distributions, and a mixture model of four Weibull distributions. Based on (3), the corresponding numbers of parameters of the two mixture models are fourteen and nineteen, respectively. However, by comparing the result in [13] where the lognormal distribution is utilized, the performance of the mixture model is inferior. One can see that for both models, the estimated Cdfs under the median and low voltage levels significantly deviate from the Kaplan-Meier estimates. However, from Fig. 4 that illustrates the shape and scale parameters in the mixture model of four Weibull distributions (the case for the mixture model of three Weibull distributions are omitted), the adopted life-stress relationships appear to be appropriate. In other words, a more complex mixture of Weibull model with more parameters must be considered if better estimates are expected.

TABLE I. ALT DATA OF A TYPE OF MINIATURE LAMP

Stress	Lifetimes in hours ("+" censored)							
5 V (z_1=1)	20.5	22.3	23.2	24.7	26	34.1	39.6	41.8
	43.6	44.9	47.7	61.6	62.1	65.5	70.8	87.8
	118.3	120.1	145.4	157.4	180.9	187.7	204	206.7
	213.9	215.2	218.7	254.1	262.6	293	304	313.7
	314.1	317.9	337.7	430.2				
3.5 V (z_2=0.5)	37.8	43.6	51.1	58.6	65.5	65.9	75.6	82.5
	88.1	89	106.6	113.1	121.1	121.5	128.3	151.8
	171.7	181	202.7	211.7	230.7	249.9	275.6	285
	296.2	358.5	379.8	434.5	493.1	506.1	570	577.7
	876.3	890+	890+	922	941+	941+		
2 V (z_3=0)	223.1	254	316.7	560.2	679	737	894.4	930.5+
	930.5+	930.5+	930.5+	930.5+	930.5+	930.5+	930.5+	930.5+
	930.5+	930.5+	930.5+	930.5+	930.5+	930.5+	930.5+	930.5+

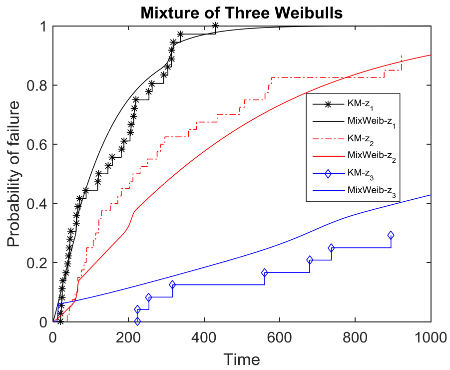

Fig. 2. Statistical fits of the mixture model of three Weibull distributions.

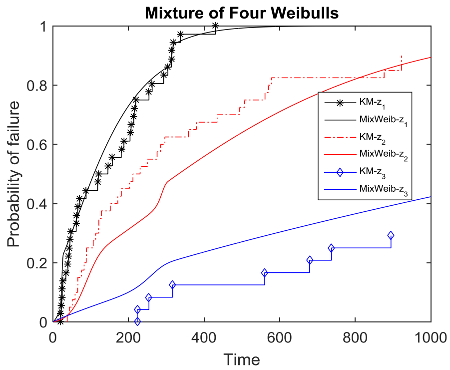

Fig. 3. Statistical fits of the mixture model of four Weibull distributions.

C. *Results of the Proposed Coxian-based General Method*

To demonstrate the superior performance of the proposed Coxian-base method with the general life-stress relationship formulation (exponential functions), models with different numbers of phases are obtained. Figs. 5 and 6 illustrate the estimation performance of the three-phase Coxian-based model and the seven-phase alternative. From (5) and (6), the numbers of model parameters of the two models are eight and twenty, respectively. One can see that the proposed method is able to provide quite adequate statistical fits. From Fig. 7, the three-phase Coxian-based model is suggested based on the AICs. For this model, the non-parametric bootstrap method is implemented, and the resulting 95% CIs are shown in Fig. 8.

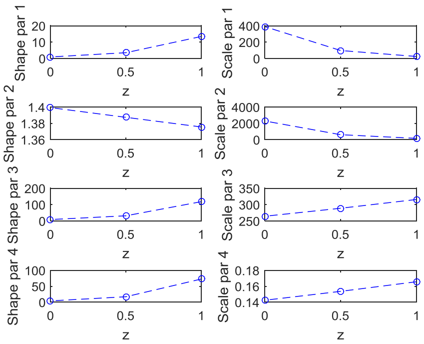

Fig. 4. Shape and scale parameters in a mixture of four Weibull distributions.

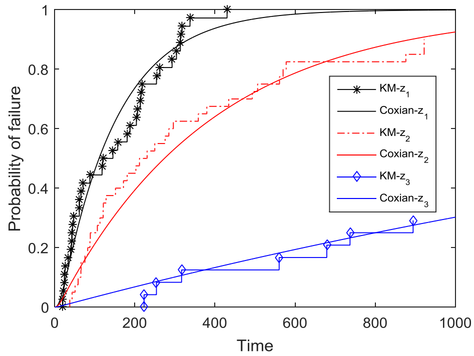

Fig. 5. Statistical fits of the three-phase Coxian model with general life-stress relationship.

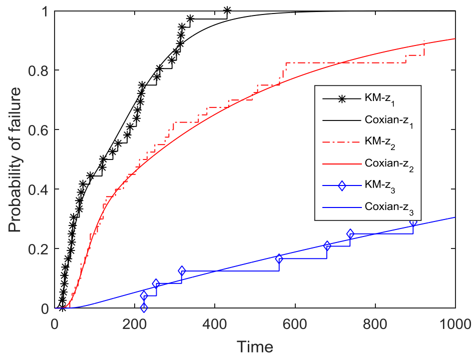

Fig. 6. Statistical fits of the seven-phase Coxian model with general life-stress relationship

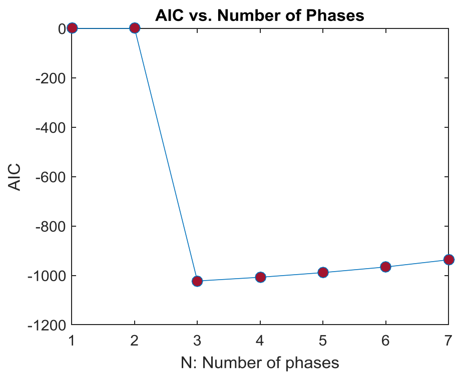

Fig. 7. AIC values for Coxian-based models with different numbers of phases.

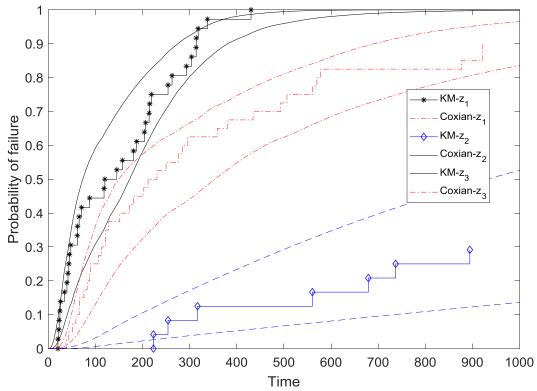

Fig. 8. 95% CIs for the Cdfs under different voltage levels using the three-phase Coxian model with general life-stress relationship.

V. CONCLUSION

This research addresses general methods for modeling lifetime data with covariates. A Coxian-based method that incorporates a flexible life-stress relationship is proposed and compared with a mixture of Weibull distributions. The advantage of the proposed general data analysis method is that an adequate fit to lifetime data can be obtained by gradually changing the number of phases of the associated CTMC. Without assuming other particular probability distributions for the lifetimes, such as extreme-value distributions, lognormal distribution, and gamma distributions, this method can well represent the underlying lifetime distribution. A maximum likelihood-based approach is developed for estimating the model parameters and determining the number of phases of the model. A non-parametric bootstrap approach is utilized to construct the interested confidence intervals. The numerical example demonstrates that the proposed general method indeed provides practitioners with a convenient statistical tool for modeling lifetime data with covariates. Compared with the mixture of Weibull distributions model, the proposed method is able to provide more accurate estimates with comparable model complexity.

ACKNOWLEDGMENT

The authors would like to thank the technical committee and conference organizers for organizing the conference. They also appreciate the three knowledgeable reviewers for providing the valuable comments that help the authors improve the quality and presentation of this paper.

REFERENCES

[1] W. Kuo, "Challenges related to reliability in nano electronics," IEEE Transactions on Reliability, vol. 55, pp. 569-570, 2006.

[2] J. Sethuraman and N.D. Singpurwalla, "Testing of hypotheses for distributions in accelerated life tests," Journal of the American Statistical Association, vol. 77, pp. 204-208, 1982.

[3] L.A. Escobar and W.Q. Meeker, "A review of accelerated test models," Statistical Science, vol. 21, pp. 552-577, 2006.

[4] E.A. Elsayed, "Overview of reliability testing," IEEE Transactions on Reliability, vol. 61, pp. 282-291, 2012.

[5] S. Asmussen, O. Nerman, and M. Olsson, "Fitting phase-type distributions via the EM algorithm," Scandinavian Journal of Statistics, vol. 23, pp. 419-441, 1996.

[6] M.A. Johnson and M.R. Taaffe, "Matching moments to phase distributions: density function shapes," Communications in Statistics: Stochastic Models, vol. 6, pp. 283-306, 1990.

[7] J.E. Ruiz-Castro, R. Rerez-Ocon, and G. Fernandez-Villodre, "Modelling a reliability system governed by discrete phase-type distributions," Reliability Engineering & System Safety, vol. 93, pp. 1650-1657, 2008.

[8] E. Jonsson, M. Andersson, and S. Asmussen, "A practical dependability measure for degradable computer systems with non-exponential degradation," In Proceedings: IFAC Falut Detection, Supervision and Safety for Technical Processes, Safeprocess'94, Espoo, Finland, pp. 231-237, 1994.

[9] H. Liao and H. Guo, "A generic method for modeling accelerated life testing data," Proceedings of the 59th Annual Reliability and Maintainability Symposium, Orlando, FL, January 28-31, 2013.

[10] V. Bagdonavičius and M. Nikulin, Accelerated Life Models - Modeling and Statistical Analysis, New York: Chapman & Hall/CRC, 2002.

[11] T. Osogami and M. Harchol-Balter, "Closed form solutions for mapping general distributions to quasiminimal PH distributions," Performance Evaluation, vol. 63, pp. 524-552, 2003.

[12] H. Liao, Z. Ye, and R. Guo, "An Erlang-Coxian-based method for modeling accelerated life testing data," in Advances in Through-Life Engineering Services, Edited by Redding, L., Roy, R. and Shaw, A., Springer International Publishing, Switzerland, 165-186, 2017.

[13] H. Liao and E.A. Elsayed, "Equivalent accelerated life testing plans for log- location-scale distributions," Naval Research Logistics, vol. 57, pp. 472-488, 2010.