SAND2006-5448C

Updating a User Friendly Combined Lifetime Failure Distribution

Daniel Briand, University of New Mexico, Sandia National Laboratories

James E. Campbell, PhD., Sandia National Laboratories

Aparna V. Huzurbazar, PhD., University of New Mexico

Key Words: consequence analysis, health management, prognostics, PHM

1 INTRODUCTION

One of the primary purposes of current enterprise level modeling efforts is to use component/system reliability estimates along with inventory levels, maintenance and inspection schedules, and operational requirements to optimize supply/repair chain processes. In addition, prognostics and health management modeling uses component/system reliability estimates as a baseline from which the data from sensors and maintenance events along with data fusion techniques can determine component health trends. These component health trends may help predict failure far enough in advance to be able to modify operations and maintenance schedules for the purposes of maximizing system availability or minimizing maintenance and spares costs. In either case, once a characterization of the component's lifecycle in terms of failure probability is established, a methodology for how to update that characterization based on the availability of new data is required, a methodology that ensures the resulting distribution is useful to the modeler.

Reliability models depend on failure distributions to characterize the probability of failure over the lifetime of a component. Many types of components typically will have a bathtub-shaped failure rate life distribution. This distribution is commonly characterized by a decreasing failure rate during the early portion of its life, a constant failure rate during the useful portion of its life, and an increasing failure rate during the wear out portion of its life, as shown in Figure 1. During the early portion of its life, failures are typically caused by manufacturing defects. During the useful portion of its life, component failures are usually caused by chance, perhaps as a result of overstress or a shock to the system. The wear out portion of its life is characterized by wear or accumulated damage that exceeds allowable limits for normal operation [1]. In many supply/repair models, the failure rate distributions for components model only the useful life period, typically with a constant failure rate that does not take into account the aging process and the wear out problems that will occur [2]. However, when modeling at the unit or enterprise level, being able to use the failure characteristics across a component's lifetime provides greater accuracy and usefulness of the model.

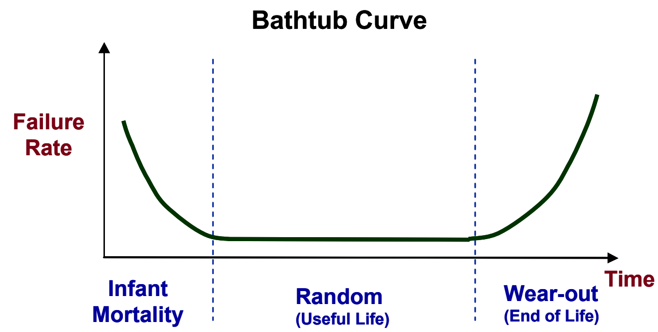

Figure 1. The Bathtub Curve[3]

Knowing how the failure rate will change over time is also important to being able to predict a failure of an individual component early enough to be able to modify operations and maintenance scheduled in order to maximize availability. This capability is commonly termed prognostics and health management (PHM). The prognostic capability relies on knowing how an individual component's failure rate deviates from the "average" or expected component's failure rate distribution. The change can be analyzed to determine or predict the component's remaining useful life. This prognostic capability relies on sensors operating in real-time and/or inspections to detect changes in a component's health, and data fusion algorithms that use that information to predict the change in time to failure or remaining useful life. Whether to optimize the supply and repair chain process or to implement an effective PHM program, the accurate portrayal of a component's failure distribution across its entire lifecycle is critical to maximizing a system's availability while minimizing parts and maintenance costs.

The goal of this ongoing research is to better understand how to correctly update time-to-failure (TTF) distributions, based initially on sparse data, data from similar components, and expert opinion, with new observations and sensor data. The remainder of this paper will discuss updating bathtub shaped TTF distributions. More specifically, the next section will briefly describe the Sandia National Laboratories' developed Combined Lifecycle (CMBL) distribution used for enterprise level and PHM component reliability representation. The subsequent section will describe three possible approaches to updating bathtub shaped TTF distributions, followed by a section that provides greater detail on initial results of the first of the approaches. Finally, the results are summarized and possible areas for further exploration are suggested.

2 COMBINED LIFECYCLE DISTRIBUTION

Early in a systems life cycle, reliability data may not abundant, as opposed to later in the systems lifecycle, where operational failure data becomes available. Where data is not abundant, expert opinion is solicited. Sometimes data from similar components can be used but must be updated with expert opinion. The expert opinion may come from engineers/technicians who are typically not statisticians or reliability experts so it helps immensely to be able to elicit the necessary information using more common terms and concepts, i.e., how long is the burn-in phase, what percent of total component failures are a result of burn-in, what is the mean life expectancy of the component given it makes it past burn-in, etc. Trying to use common failure distributions, such as combinations of the Gamma and/or Weibull distributions, may be quite involved since converting expert opinion to a parameter value would most likely require several iterative steps [4]. Despite considerable published works in bathtub shaped failure distributions, few practical models are available [5].

To help simplify the component failure characterization process early in a system's life cycle, Sandia National Laboratories is currently using the CMBL distribution for enterprise level and PHM component reliability representation. The CMBL distribution assumes a linearly declining failure rate during infant mortality, a constant failure rate during normal life, and a normally distributed TTF as the component nears its end of life (Figure2). This failure distribution represents the entire component's lifetime with parameters that make it relatively easy to elicit the probability of failure distribution from subject matter experts and limited data. These parameters are:

1. The mean of the normally distributed portion of the TTF distribution.
2. The standard deviation of the normally distributed portion of the TTF distribution.
3. The probability that the component will fail during burn-in.
4. The duration of the burn-in portion of the distribution.
5. The probability that failure will occur randomly after burn-in.

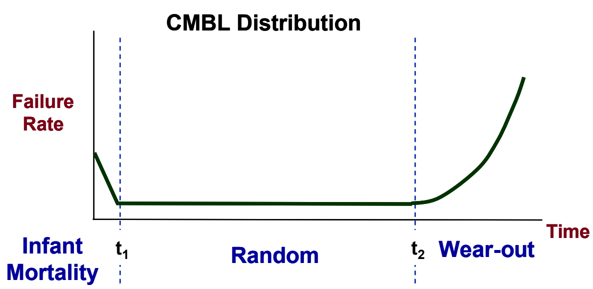

Figure 2. Combined Lifecycle (CMBL) Distribution

The explicit form of the distribution is:

$$f(t) = \begin{cases} \lambda_d e^{-\lambda_d t} & 0 \le t \le t_1 \\ \lambda_c e^{-\lambda_c t} & t_1 \le t \le t_2 \\ \dfrac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(t-u)^2} & t_2 \le t \le \infty \end{cases} \quad (1)$$

where

μ = mean of the normally distributed portion of the TTF distribution.

σ = standard deviation of the normally distributed portion of the TTF distribution.

λ_d= failure rate for the linearly decreasing failure rate portion = $(mt+b)$.

λ_c = failure rate for the constant failure rate portion

t_1= burn-in duration.

t_2 = transition from constant failure rate to the normal TTF portion.

F_1= fraction of failures occurring in the infant mortality portion.

F_2= fraction of failures occurring in the random failure portion.

While not the focus of this paper, an iteration scheme is set up to solve for λ_c and t_2 based on the ensuring that the transitions from the linearly decreasing failure rate to the constant failure rate and from the constant failure rate to the normal TTF portion are continuous [6].

Once the CMBL distribution is characterized using scarce data, data from similar components and expert opinion, the next step is to update it as new data from the components becomes available. As the system being modeled undergoes extensive testing and is used under normal operations, data on the number of failures, mean time to failure (MTTF), mean time between failures (MTBF), mean time to repair (MTTR), etc., becomes available. It is appropriate throughout the modeling process to update the parameters of the original CMBL distribution in some fashion to improve its accuracy. Updating the distribution should occur throughout a component's lifecycle since improvements in the component and changes in its use may alter its inherent reliability.

In addition, the resulting updated failure distribution must be in a usable format for the supply/repair chain or PHM model for which it resides since the resulting distribution will be fed back into the supply/repair chain or PHM model. An empirical distribution is possible, but it would be better to have the updated distribution be the same as the originating distribution. However, in some models it may not make much difference as long as the resulting distribution can be easily updated as new data becomes available. Being able to transition the CMBL distribution to a more common form such as a single commonly used distribution would most likely not work. Transition from the CMBL distribution to a bathtub distribution modeled by two or more distributions such as 3 Weibulls (one for each section) would present even greater challenges.

3 METHODOLOGY FOR UPDATING

We are investigating three different methods for updating piecewise continuous failure distributions.

3.1 *Method 1*

With the piecewise continuous CMBL distribution, treat each section, infant mortality, random, and wear-out phases, separately. Using the appropriate prior depending upon where the new data occurred, i.e., 0 to t_1, t_1 to t_2, and t_2 to infinity, assume that the new data has the same underlying distribution as the section it occurred in and use a Bayesian updating methodology to determine the posterior distribution. This posterior distribution, or an estimate of the posterior distribution, will be used as the subsequent prior as new data occurs in that section. For example, use a Gamma (α,β) as the prior distribution in the random (constant failure rate) section of the CMBL distribution and assume the new data follows an exponential distribution with a constant failure rate. From the resulting posterior distribution and a modified iterative scheme used to determine the constant failure rate λ_c, determine the new parameters of the CMBL distribution.

This method may have a problem since the new data is assumed to have the same distribution as where it falls in the current section of the bathtub curve. If the original CMBL distribution is quite different than the distribution described by the new data, a major concern would be that by presupposing that the new data should be modeled by the distribution of the section for which it occurs in the original CMBL distribution, the resulting distribution would never converge to the appropriate distribution of the new data over time. Also, reverse engineering of the iteration procedure of the CMBL distribution will be necessary to determine the five input parameters.

3.2 *Method 2*

Evaluate updating the piecewise continuous CMBL distribution as a single distribution with new data represented by different types of common distributions. Next, use a Markov chain simulation method to derive the posterior distribution. Use regression on the resulting distribution to obtain the three updated input parameters for the infant mortality and random sections of the CMBL distribution and a curve fitting technique to get the two updated input parameters for the normal section. Finally, evaluate the appropriateness of the common distributions used to represent the data and how well the regression and curve fitting techniques worked to represent the posterior distribution.

This method may present the greatest challenges since trying to determine an appropriate distribution for the new data may be difficult and may be different for different sections of the prior CMBL distribution. Determining the five input parameters may require considerable reverse engineering of the iteration procedure to ultimately obtain the input parameters from the regression and curve fitting results.

3.3 *Method 3*

Use either of the Methods 1 or 2 above to determine the posterior empirical distribution. Use this empirical distribution "as is" assuming the logistic/PHM model has the capability to handle empirical distributions. Subsequent updates using typical Markov chain simulation methods to derive the posterior distribution would use the empirical distribution as the prior distribution to arrive at a posterior as new data becomes available.

This method may be the easiest but using an empirical distribution usually increases data storage requirements, although this is usually not a significant problem in large scale supply/repair chain or PHM models. As an alternative, updating of the distributions could occur in a subroutine outside the supply/repair chain or PHM models. Probably the biggest downside to this approach is that the user does not get a feel for what a component's failure distribution looks like without some additional statistical analyses.

4 IMPLEMENTATION OF METHOD 1

This method determined if the CMBL distribution could converge to the distribution of the data through a section by section Bayesian updating methodology. In this approach, each new data point was assumed to have the same underlying failure distribution as the section of the CMBL distribution where it occurred. Several additional simplifying assumptions were also made. Only the random and wear-out portions of the bathtub curve were evaluated specifically, although the infant mortality portion was included in the process to determine if there would be any unusual behavior in the transition region around t_1. It was assumed that with a Bayesian updating methodology, the CMBL distribution would be updated one data point at a time. Since the new data created from different input parameters can be quite extreme, especially with regard to the Exponential distribution, a weighting scheme was introduced to "smooth" the convergence process. Finally, the standard deviation for the data distribution in the wear-out portion of the distribution was assumed to be known and held constant. The impact of each of these assumptions will be investigated in future work. The validation of this approach is broken down into several steps as outlined in the following paragraphs.

4.1 *Step 1*

The first step was to evaluate the process for updating the CMBL distribution with data falling within the wear-out section of the distribution. This was a relatively straightforward process. In a Bayesian updating methodology, the conjugate prior is a Normal (u, σ). Using a Normal $(u_{CMBL}, \sigma_{CMBL})$ for the prior and a $N(u_0, \sigma)$ for the single new data point results in a $Normal(u_1, \sigma_1)$ where

$$u_1 = \frac{\dfrac{u_{CMBL}}{\sigma_{CMBL}^{\;2}} + \dfrac{t_1}{\sigma^2}}{\dfrac{1}{\sigma_{CMBL}^{\;2}} + \dfrac{1}{\sigma^2}} \qquad (2)$$

This posterior Normal (u_1, σ_1) distribution becomes the next prior distribution, and this updating process is repeated for each new data point. As expected, this process results in the Normal (u_{CMBL}, σ_{CMBL}) converging to the Normal (u_d, σ) reasonably well.

4.2 *Step 2*

The second step was to evaluate the Bayesian updating process for the data points that fall within the random failure section of the bathtub distribution. The only user input available for this section is F_2, the fraction of failures occurring in the random failure portion, but the resulting λ_c, the failure rate for the constant failure rate portion, is determined within the iterative process of the CMBL distribution. Since the data occurring in the random section is being modeled by an Exponential (λ) where λ is considered a random variable, a typical conjugate prior used in many reliability applications is the Gamma (α, β). Since the CMBL estimate is being taken as a single instance of failure, the Bayesian prior is the Gamma (1, β), which is essentially an Exponential (β) as given by:

$$g(\lambda \mid \alpha, \beta) = \frac{1}{\Gamma(\alpha)\beta^{\alpha}} \lambda^{\alpha-1} e^{-\lambda/\beta} \quad (3)$$

$$where\ \alpha = 1 \quad => g(\lambda : 1, \beta) = \frac{1}{\beta} e^{-\lambda/\beta} \quad (4)$$

so $\lambda \sim Expon(\beta)$. This results in a posterior distribution that is:

$$g(\lambda \mid t, \alpha, \beta) = \frac{\frac{1}{\Gamma(\alpha)\beta^{\alpha}} \lambda^{\alpha-1} e^{-\lambda/\beta} \lambda e^{-\lambda t}}{\int \frac{1}{\Gamma(\alpha)\beta^{\alpha}} \lambda^{\alpha-1} e^{-\lambda/\beta} \lambda e^{-\lambda t}} \quad (5)$$

which becomes:

$$= \frac{1}{\Gamma(\alpha+1)(\frac{\beta}{\beta t+1})^{\alpha+1}} \lambda^{\alpha+1-1} e^{-\lambda/\beta} \lambda e^{-\lambda(t+1/\beta)}, \quad 0 < \lambda < \infty \quad (6)$$

with an expected value of:

$$E(\lambda \mid t, \alpha, \beta) = \frac{\beta(\alpha+1)}{\beta t+1} [7] \quad (7)$$

In this case, assuming successive single data point Bayesian updating, the result is a Gamma (2, $\beta_0/\beta_0 t_1+1$). The resulting $E_{Gamma}(\lambda \mid t, 2, \beta)$ is the new estimate for λ_c, which is used in the CMBL iterative process to calculate a new F_2. This change in F_2 is then used with the other unchanged parameters to the CMBL distribution, which now becomes the prior distribution for the next data update. Using this estimate alone was not was not good enough when using the successive single data point updating methodology. This result became apparent when the data represented the original distribution derived from the input parameters but the sequential updating process resulted in a mean that was not reasonably close to the expected mean. When Equation (7) was multiplied by a factor of approximately 0.64 and data represented the original distribution derived from the input parameters, the deviation from the expected mean essentially disappeared. Further evaluation of the 0.64 factor is necessary to determine any underlying theoretical implications.

4.3 *Step 3*

The third step examined updating both the random and wear-out portions of the CMBL distribution simultaneously. The primary goal was to determine if the resulting Bayesian updating scheme for both sections simultaneously resulted in a convergence to the distribution of the data. Particular attention was paid to the transition region around t_2 since this requires an iterative scheme to ensure the failure distribution of the two sections remain continuous.

Several evaluation runs that included all three sections of the CMBL distribution were made to determine the characteristics of convergence to the distribution of the data. As an example, the Baseline, Normal, and Random CMBL distribution sets of parameters are shown in Table 1. The baseline distribution provides the starting or initial prior distribution. Using data from the Normal set of parameters, which changes μ only, focuses the convergence on updating the normal portion of the CMBL distribution although the random portion has to adjust as well. Using data generated from the Random set of parameters, which again changes μ only, focuses the convergence on updating the random portion of the CMBL distribution although the normal portion has to adjust appropriately.

	Baseline	Normal	Random
μ	200	250	150
σ	20	20	20
F_1	.1	.1	.1
F_2	.2	.2	.2
t_1	10	10	10

Table 1. Baseline, Normal, & Random Input Parameters

Starting with the baseline input parameters for the CMBL distribution, successive CMBL distributions with their updated parameters were created using the new data (failure times) from a CMBL distribution with the normal input parameters. A weighting scheme that essentially provided a user defined fraction of the change between the old mean and the new data point tended to minimize extreme new data values but tended to increase the number of iterations to convergence. Convergence to the distribution of the data generated from the normal input parameters was relatively consistent and quick as shown in Figure 1.

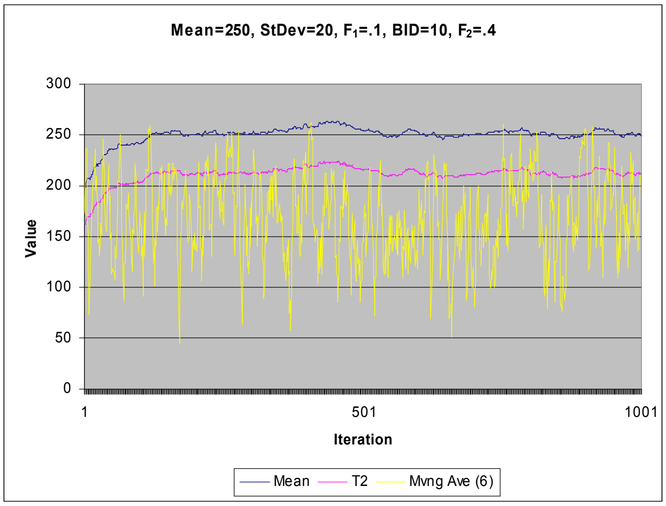

Figure 1. Convergence to Normal CMBL Data

In converging from a mean of 200 to a mean of 250, the t_2 also converged to 212, approximately the anticipated value of 211 from the starting value of 164. Convergence appears to occur within about 200 iterations. A moving average of the data, averaged over six data points, shows the variability in the data where the data spans all portions of the bathtub distribution; infant mortality, random, and wear-out.

Convergence to the distribution of the data generated from the random input parameters was relatively consistent but not as quick as shown in Figure 2. It took about 250 iterations to reach near the new mean of 150. The t_2 converged to the anticipated value of 117. Again, a moving average of the data, averaged over six data points, shows the variability in the data where the data spans all portions of the bathtub distribution.

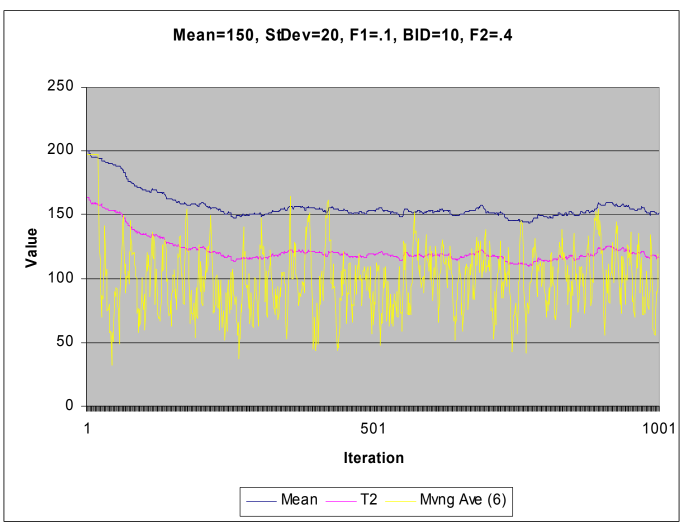

Figure 2. Convergence to Random CMBL Data

A comparison of the results of the baseline, normal, and random parameters is shown in Figure 3. The comparison shows an appropriate shift in the mean of the distribution while an appropriate shift in λ_c occurs to accommodate the shift. For example, when the mean of the distribution of the data causes a shift from 200 to 150, the failure rate λ_c of the random section of the CMBL distribution increases to ensure F_2 remains at constant at 0.4. This increase in λ_c results in the shift in the line in the random portion of the CMBL distribution. The opposite occurs when the mean of the distribution of the data shifts from 200 to 250, as expected.

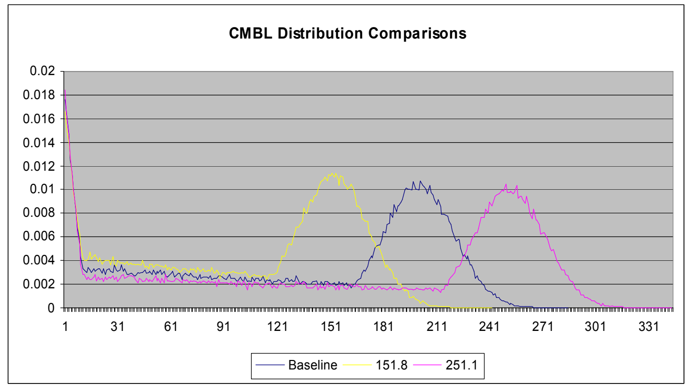

Figure 3. Baseline, Random, and Normal Comparisons

The next step in the evaluation of Method 1 will be to hold the mean constant and vary F_2, the fraction of failures occurring in the random failure portion. This step may require an iteration scheme that finds t_2 while holding λ_c constant whereas normally, λ_c and t_2 varies in the CMBL iteration scheme. Once this is successfully accomplished, updating the baseline distribution to the data distribution by varying several different combinations of the input parameters will follow. Finally, an evaluation of all three sections, infant mortality, random, and wear out, should follow.

SUMMARY & CONCLUSIONS

The results obtained in evaluating Method 1 show initial promise in finding an acceptable method for updating the CMBL distribution one data point at a time. In addition to completing the research outlined in the previous paragraph, additional research must be accomplished to determine if updating the CMBL distribution with sets of new data (instead of one data point at a time) will provide the same or perhaps better/quicker results. Evaluation of Methods 2 and 3 should also proved greater flexibility in the use of the CMBL and with some generalizing modifications, this should allow application across other sectional TTF failure models.

The method developed in this effort for updating the CMBL distribution and other TTF distributions, may be extremely valuable in enhancing maintenance planning and real-time situational awareness processes. This method, used in enterprise level and PHM modeling, should more accurately help provide instant feedback on the current status of equipment; provide tactical assessment of the readiness of equipment for the next campaign; identify parts, services, etc. that are likely to be required during the next campaign; provide a realistic basis for scheduling and optimizing equipment maintenance schedules; and help ensure that the useful life of expensive components is taken full advantage of while reducing the incidence of unplanned maintenance.

REFERENCES

1. Wolstenholme, Linda C. Reliability Modelling, A Statistical Approach. Chapman & Hall/CRC, New York, 1999.
2. Caruso, Hank. Too Much Time in the Bathtub (Curve) – An Aging Aircraft Paradigm that Doesn't Hold Water. Joint Council on Aging Aircraft, Aging Aircraft Conference Technical Paper, 2005.
3. Wilkins, Dennis J. The Bathtub Curve and Product Failure Behavior Part Two – Normal Life and Wear-Out, *Reliability Hot Wire*, Issue 22, December 2002.
4. Booker, Jane M., et al. PREDICT: A New Approach To Product Development and Lifetime Assessment Using Information Integration Technology. Submitted to *Handbook of Statistics: Statistics in Industry*, September, 2000.
5. Xie, M. & Lai, C. D. Reliability Analysis Using an Additive Weibull Model with Bathtub-Shaped Failure Rate Function. *Reliability Engineering and System Safety*, Vol 52, (1995), 87-93.
6. Campbell, James E. The Wearout Distribution – Revised. Unpublished paper, Sandia National Laboratories, 27 January 2002.
7. Martz, Harry F. and Waller, Ray A. Bayesian Reliability Analysis. John Wiley & Sons, New York, 1982.

BIOGRAPHIES

Daniel Briand
P.O. Box 5800 MS 1011
Sandia National Laboratories
Albuquerque, NM 87185-1011
Sandia National Laboratories

e-mail: dbriand@sandia.gov

Daniel Briand is a Senior Member of the Technical Staff at Sandia National Labs focusing on reliability and prognostic issues for system sustainment and readiness programs. His background is in operations analysis and aircraft operations & training, the majority of which stems from having served 20 years in the US Air Force as an operations research analyst and helicopter pilot. Dan is a graduate of the US Air Force Academy, has a MS in Operations Research from the Air Force Institute of Technology, and is currently pursuing a Ph.D. in Statistics from the University of New Mexico

James E. Campbell, Ph. D.
P.O. Box 5800 MS 1011
Sandia National Laboratories
Albuquerque, NM 87185-1011
Sandia National Laboratories

e-mail: jecampb@sandia.gov

James Campbell is a Distinguished Member of the Technical Staff at Sandia National Labs. Prior to joining Sandia, he was a Vice President at Intera Inc. in Austin, Texas. His areas of expertise include reliability analysis, decision analysis, analysis of complex systems of systems, and software design and engineering. He has bachelor's degrees in mathematics and physics and a Ph.D. in physics.

Aparna V. Huzurbazar, Ph.D.
Department of Mathematics and Statistics
MSC03 2150, 1 University of New Mexico
Albuquerque, NM 87131-0001
University of New Mexico

e-mail: aparna@stat.unm.edu

Aparna V. Huzurbazar is Associate Professor of Statistics (tenured) at the University of New Mexico. Her background is in aerospace engineering and she is the author of numerous technical articles in reliability and survival analysis, Bayesian statistics, stochastic processes, and applications to engineering systems. She is the author of *Flowgraph Models for Multistate Time-to-Event Data* published by Wiley in 2005. She holds a BS in Aerospace Engineering from the University of Colorado at Boulder, a BA in Mathematics from Claremont McKenna College, and a PhD in Statistics from Colorado State University. She holds concurrent appointments as adjunct staff at the RAND Corporation in Santa Monica, CA, and faculty affiliate at Los Alamos National Laboratory, Los Alamos, NM.