

# Bayesian Hierarchical Models for Common Components across Multiple System Configurations

Kassandra M. Fronczyk  
Institute for Defense Analyses  
Alexandria, VA  
kfronczyk@ida.org

Rebecca M. Dickinson  
Institute for Defense Analyses  
Alexandria, VA  
rdickins@ida.org

James F. Gilman  
North Carolina State University  
Raleigh, NC  
jfgilman@ncsu.edu

Alyson G. Wilson  
North Carolina State University  
Raleigh, NC  
agwilso2@ncsu.edu

Laura J. Freeman  
Institute for Defense Analyses  
Alexandria, VA  
lfreeman@ida.org

June 28, 2017

## Abstract

In this paper, we use a Bayesian hierarchical model to assess the reliability of the Joint Light Tactical Vehicle (JLTV), which is a family of vehicles. The proposed model effectively combines information across three phases of testing and across common vehicle components. The analysis yields estimates of failure rates across failure modes and vehicles as well as an overall estimate of the failure rate for the family of vehicles. We are also able to obtain estimates of the Fix Effectiveness Factor (FEF),

a measure of how well repairs between test phases improve failure rates. In addition to using all data to improve on current assessments of reliability and reliability growth, we illustrate how to leverage the information learned from the three phases to determine the length of testing needed to demonstrate a given reliability threshold in subsequent testing.

*Keywords:* Assurance Testing, Combining Information, Defense Acquisition, Reliability, Reliability Growth

# 1 Introduction

Reliability is a high priority in the testing of military systems. A common interest is *reliability growth*, which tracks the change, and ideally improvement, in the reliability of a system as it moves through testing phases and undergoes periodic corrective actions. Ideally, plans for future test events should be based on what is observed in previous testing.

In this paper, we look at combining information across multiple test phases with the objective of planning a future test. When combining information, there is no omnibus solution. Rather, models need to be carefully considered and evaluated to ensure that they accurately reflect the data and the underlying physical processes. Our analysis is motivated by reliability data from the Joint Light Tactical Vehicle (JLTV), which is a family of vehicles designed to replace one-third of the legacy high-mobility multipurpose wheeled vehicle (Humvee) fleet. There are four unique types of vehicles, and within types there are many possible variants. The primary mission of the family of vehicles is to provide ground mobility that is deployable worldwide and capable of operating across the range of military roles including combat, sustainment, police action, peace-keeping, and security patrol, in all weather and terrain conditions.

The JLTV has been through a series of test events as it has been developed. Engineering and Manufacturing Development (EMD) included three phases of testing, where fixes occurred only during a set Corrective Action Period (CAP) between each phase. For every vehicle, each failure encountered during testing was recorded and attributed to a specific failure mode. Failures discovered during mission execution that result in an abort or termination of a mission in progress are scored as Operational Mission Failures (OMF), and failures of mission essential components are scored as Essential Function Failures (EFF). The EFFs tend to include a large portion of the failure modes that drive maintenance costs and reduce system availability. While requirements are typically written in terms of OMFs, because OMFs are by definition EFFs that occurred at critical times, combining the failures provides a more robust reliability estimate. Since fixes were delayed to CAPs, analysis generally shows a distinct jump in the system reliability between test phases.

We propose a model that can effectively combine information across the three phases of testing and across common vehicle components. We use a Bayesian hierarchical model to assess the reliability of the family of vehicles. The model accounts for the commonality across vehicle types and allows for uncertainty quantification. Inferential objectives from the proposed model include an estimate of the mean miles between failure (MMBF) for each vehicle, failure mode, and phase of test, and an assessment of the effectiveness of the corrections completed in the CAPs. Additionally, because future tests are being planned,

we propose methodology to leverage the information learned from the three phases to determine the length of testing needed to demonstrate a given reliability threshold in subsequent testing.

## 2 Data

Eight vehicle prototypes were used for reliability, availability, and maintainability (RAM) testing. There were four two-seat utility vehicles, two four-seat general purpose vehicles, a two-seat heavy guns carrier, and a two-seat close combat weapons carrier. The close combat weapons carrier, one general purpose, and two utility vehicles were tested in Aberdeen, Maryland, where it is much colder in the winter months but the terrain is not as harsh, and the heavy weapons carrier, one general purpose, and two utility vehicles were tested in Yuma, Arizona, where it is warmer but with more challenging terrain, to attain testing in different environmental and terrain conditions. The vehicles were driven roughly 3000 to about 5000 miles in Phase 1, between 4500 and 7500 miles in Phase 2, and between 5000 and 6700 miles in Phase 3.<sup>1</sup>

There are many similarities between vehicle types; in fact, the most dissimilar vehicles have over 80% common parts. Consequently, some failure modes, such as brakes and radios, will be common to all eight vehicles. Other failure modes will be related, but not identical, as with hydraulics, frame, and body.

There are 91 OMFs and 1321 EFFs attributed to 26 failure modes recorded across all eight vehicles and all three test phases. In any given phase of test, every vehicle is not guaranteed to have a failure of all 26 failure modes, and each test phase does not necessarily end with a failure. Therefore, we have 624 right censored observations, accounting for miles driven without observed failures.

## 3 Methodology

A standard reliability analysis employed by the Department of Defense (DoD) test community considers each test phase independently and uses the exponential distribution to model the miles between failure [1]. The traditional analysis is overly simplistic and ignores valuable information learned about the individual vehicles and their failure modes. Consider the following alternative approach.

---

<sup>1</sup>The exact mileage has been altered to protect proprietary information.

In Test Phase 1, we assume the vehicle miles at failure,  $y$ , follow an exponential distribution with a failure rate parameter  $\lambda_{ij}$ . Introducing notation,

$$y_{ijk} \mid \lambda_{ij} \sim \text{Exp}(\lambda_{ij}), \quad i = 1, 2, \dots, v \quad j = 1, 2, \dots, s \quad k = 1, 2, \dots, n_{ij} \quad (1)$$

where  $y_{ijk}$  are the miles between failure for vehicle  $i$  failure mode  $j$ ,  $v$  is the number of vehicles,  $s$  is the number of failure modes, and  $n_{ij}$  are the number of failures of vehicle  $i$  failure mode  $j$ . The number of failure modes is assumed fixed and known *a priori*.

The prior distribution on the exponential failure rate parameter,  $\lambda_{ij}$ , depends on whether failure mode  $j$  is considered to be common across vehicles or related but not identical. For the related failure modes, we place a gamma prior distribution on the collection of  $\lambda_{ij}$ ; in other words, we assume each vehicle has a distinct failure rate in failure mode  $j$  but they arise from a common gamma distribution.

If failure mode  $j$  is considered common across vehicles, the collection of failure rates is collapsed to a single parameter,  $\lambda_{ij} = \lambda_j$ . As with the related failure modes, a gamma prior distribution is placed on the single failure rate. The prior distributions are independent across failure mode, and can potentially have different hyperparameters.

The Phase 1 analysis yields an estimate of the failure rate  $\lambda_{ij}$  for each of the vehicles for failure modes that are related. We are assuming the vehicles are conditionally independent, therefore the failure rate estimate for the family of vehicles for such failure modes can be found by  $\sum_i \lambda_{ij}$ . For failure modes that are common across vehicles the Phase 1 analysis yields a  $\lambda_j$ , which is the failure rate for the family of vehicles. The overall failure rate across all failure modes can be found by  $\sum_j \lambda_j$ .

After the first CAP, Test Phase 2 begins with the repaired vehicles. To capture these revisions, the PM2 reliability growth model [7] is often used. This model explicitly captures testing phases, choices about which failure modes to correct, and the potential of not completely eliminating a failure upon repair. One of the downsides of PM2 is that many parameters of potential interest, such as the Fix Effectiveness Factor (FEF), which measures how much repairs improve failure rates, are typically fixed. A common value for FEF is 0.70. We follow the premise of this type of model, but allow a more data-driven result that is less dependent on hard-coded assumptions.

For Phase 2 data, we assume the failure rates are non-decreasing; either the fixes were effective or had no effect, but did not degrade the family of vehicles. Using this assumption, we write the Phase 2 rate parameters as a function of the rate parameters found in Phase 1. In particular, we define  $\lambda_{ij}^{P2} = (1 - \rho_j)\lambda_{ij}^{P1}$  where  $\rho_j$  is the FEF for failure mode  $j$ . Given this definition of  $\lambda_{ij}^{P2}$ , we again model the failure miles for a given vehicle and failure mode

using the exponential distribution. We assume the prior distribution for the FEF is a beta distribution.

After Phase 2 we can look again at failure rates across failure modes and vehicles and obtain an overall estimate of the rate for the family of vehicles. The FEF provides an estimate of how effective the first CAP was and which failure modes are persisting.

The analysis of Test Phase 3 follows the same pattern as that shown in Phase 2. At the end of Phase 3, we can look at failure rates across failure modes and vehicles and obtain an overall estimate of the rate for the family of vehicles. Future tests are planned based on the inferences of Phase 3.

## 4 Discussion

Preliminary results provide indications that our model is successfully capturing the failure rates across failure modes, inferring logical fix effectiveness factor distributions, and providing reliability estimates that can be leveraged for future test planning. In the traditional analysis, we do not typically look at failure rates of the failure modes, nor would we be able to leverage the commonalities across vehicles. The proposed modeling framework overcomes these limitations and allows for more realistic reliability estimates for vehicles with no observed failures. Additionally, our data-driven approach for reliability growth improves on the current growth models being used in the DoD, which are based solely on fixed quantities set by management.

While our basic approach provides reasonable initial inferences, there are several extensions to incorporate. For example, reliability estimates for vehicles with no observed failures are sensible; however, we should leverage more information across vehicles. Because the correlations induced by our current model are fairly low, potential methods for inducing more correlation between the vehicles in the related failure modes are being investigated. We also need to look more closely at our assumption of a fixed number of failure modes. Currently, the model specification assumes the number of failure modes is fixed at 26, as this is the number of unique failure modes that were observed in the data. It is possible, though, that in a future test phase a new failure mode will be discovered. To account for this, the number of failure modes may need to be specified as a random quantity in the model with a prior distribution, or we may look at limiting behavior as the number of failure modes becomes large.

Finally, inferences learned from the first three phases of testing will be used to develop a test plan for the next phase. Here the objective is to demonstrate that at a desired

level of confidence, the system will meet or exceed a specified requirement. The methods to be employed will be similar to the assurance testing as discussed by Hamada et al. [2]. Bayesian assurance tests are used to insure that the reliability of an item meets or exceeds a specified requirement with a desired probability. Although practitioners often use “assure” and “demonstrate” synonymously, Meeker and Escobar [3] distinguish between reliability demonstration and reliability assurance testing. A *reliability demonstration test* is essentially a classical hypothesis test, which uses only the data from the test to assess whether the reliability-related quantity of interest meets or exceeds the requirement. A *reliability assurance test*, however, uses supplementary data and information to reduce the required amount of testing. Bayesian reliability assurance tests [4, 5, 6, 7] have not addressed either analysis in a reliability growth context, nor the incorporation of covariates.

## References

- [1] Director, Defense Test and Evaluation. Test and evaluation of system reliability availability maintainability - A primer; Third Edition (1982).
- [2] M. S. Hamada, A. G. Wilson, C. S. Reese and H. F. Martz. Bayesian Reliability. Springer (2008).
- [3] W. Q. Meeker and L. A. Escobar. Reliability: the other dimension of quality. *Quality Technology and Quantitative Management* 1, 1-25 (2004).
- [4] M. S. Hamada, A. G. Wilson, B. P. Weaver, R. W. Griffiths and H. F. Martz. Bayesian binomial assurance tests for system reliability using component data. *Journal of Quality Technology* 46, 24-32 (2014).
- [5] V. E. Johnson, T. L. Graves, M. S. Hamada, and C. S. Reese. A hierarchical model for estimating the reliability of complex systems. In Bayesian Statistics, Vol. 7, J. M. Bernardo, M. J. Bayarri, J. O. Berger, A. P. Dawid, D. Heckerman, A. F. M. Smith, and M. West eds., Oxford, UK: Oxford University Press (2003).
- [6] M. S. Hamada, H. F. Martz, C. S. Reese, T. L. Graves, V. E. Johnson, and A. G. Wilson. A fully Bayesian approach for combining multilevel failure information in fault tree quantification and optimal follow-on resource allocation. *Reliability Engineering and System Safety* 86, 397-405 (2004).

- [7] P. M. Ellner and J. B. Hall, An approach to reliability growth planning based on failure mode discovery and correction using AMSAA projection methodology. *IEEE Proceedings of the Annual Reliability and Maintainability Symposium* (2006).