Predicting the Reliability of a Complex Military System

Carol Vesier, U. S. Army Test & Evaluation Command
Angelo Christino, U. S. Army Test & Evaluation Command
Brian Hampton, U. S. Army Test & Evaluation Command

Key Words: reliability growth, reliability projection

SUMMARY & CONCLUSIONS

The purpose of this paper is to discuss the reliability projection for a complex military system. For the purposes of this paper we will consider this a mobile system with standalone subsystems integrated onto the platform. The reliability of the overall system is determined by these subsystems as well as the reliability of the mobile platform itself. At the time the reliability projection was made, the system had been previously tested in developmental and operational environments. Following the first developmental test (DT), the system entered into operational testing (OT). After completion of the initial DT and OT testing, the system underwent corrective actions to address issues uncovered in testing. After correction, the system underwent additional developmental testing. After the last phase of developmental testing, analysis was done to project the system’s reliability prior to entering the second OT. While taking into consideration the subsystems and their impact on the overall affect on the platforms reliability, the analysis concluded the reliability of a specific component greatly affected the projected reliability of the platform in an operational environment.

BACKGROUND

The Army commits considerable financial resources to the testing and evaluation of systems. There are two types of test events: Developmental and Operational. From the reliability perspective, the two types of tests serve different purposes.

- Developmental testing typically is longer duration and performed in a realistic physical test environment by professional testers. The purpose of this testing is two-fold: identify failures and assess corrective actions.
- Operational testing is typically shorter duration and operationally realistic. Part of operational realism is to utilize Soldiers in the test. Another aspect of operational realism is that the test is not as tightly controlled. A key purpose of this test is to evaluate the impact of realism on system reliability.

Key in the test and evaluation process is projecting reliability in the next test phase. The projected reliability is used to assess risk that a system will fall short of reliability thresholds. In this case, the threshold reliability for the system was 1000 Mean Miles between Failures (MMBF). Depending on the perceived risk, future tests may be delayed, canceled, or modified. For this system, reliability varied widely between tests, as shown in Figure 1. Figure 1 shows the reliability point estimate as well as the 80% upper and lower confidence bounds. Examining Figure 1 highlights issues in projecting the reliability in OT-II, the next test phase.

- What caused the large reliability degradation between DT-I and OT-I? Do we expect this drop to reoccur in OT-II?
- What caused the large reliability improvement in DT-II? Do we expect this improvement to be maintained in OT-II?

UNDERSTANDING RELIABILITY TEST HISTORY

2.1 Identifying the root cause of the OT-I reliability drop

Ideally, a well-designed system’s reliability will not be affected by the operator or environment, assuming the operator has been trained to use the system. Reality is that tested reliability typically drops when you add operational realism. Normally, one would expect to see reliability degradation around 25%, but there is large variability around this number. For this system, the reliability degradation
associated with adding operational realism to the mix was very high, 56%. Figure 2 shows that Component A could explain all of the reliability degradation. In this case, the integration of Component A on the System increased the likelihood of failure due to damaging Component A.

2.2 Understanding the root cause of reliability improvement in DT-II

Part of the Army development process is to evaluate failures seen during testing and decide whether or not the failure should be corrected through Corrective Actions (CA). Failure modes that are not corrected are called A-modes. Those that are corrected are called B-Modes. Figure 3 shows that none of the corrected failure modes reoccurred in DT-II. The zero reoccurrence completely accounts for the dramatic reliability improvement seen in DT-II.

3 PROJECTING RELIABILITY IN NEXT TEST PHASE

We used the following equations to project the reliability of the next test phase.

\[
\text{MTBF}_{\text{Projected,DT}} = \frac{1}{\lambda_{\text{A}} + \lambda_{\text{B, unseen}} + \lambda_{\text{B, seen}} (1 - \text{FEF})} \tag{1}
\]

\[
\text{MTBF}_{\text{Projected,OT}} = \text{MTBF}_{\text{Projected,DT}} \times (1 - \Delta_{\text{DT to OT}}) \tag{2}
\]

where:
- \(\lambda_{\text{B, seen}}\) is the failure rate of seen B-Modes
- \(\lambda_{\text{B, unseen}}\) is the failure rate of unseen B-Modes
- \(\lambda_{\text{A}}\) is the failure rate of A-Modes
- FEF is the aggregate Fix Effectiveness Factor
- \(\text{MTBF}_{\text{Projected,DT}}\) is the projected DT reliability
- \(\Delta_{\text{DT to OT}}\) is the degradation factor associated with going from a DT to an OT environment.

To evaluate the above equations required leveraging what we learned from each test phase to project the future reliability. Because of the statistical dissimilarity between the results from each test phase, we could not directly combine the results. Instead we decided to use the data from each phase as follows.
- DT-I, the initial and longest test phase, would be used to establish a failure rate baseline \(\lambda_{\text{B, seen}}\) prior to accounting for the impact of corrective actions and operational realism (Soldiers/Environment).
- Comparison of DT-I to DT-II would be used to estimate the impact of corrective actions on reliability.
- The occurrence rate of new B-Mode failures \(\lambda_{\text{B, unseen}}\) would be estimated using AMSAA Maturity Projection Model (AMPM). AMPM predictions would be validated by comparing the predictions for DT-II to the actual number of new B-Modes uncovered in DT-II.
- The occurrence rate of A-Modes \(\lambda_{\text{A}}\) would be calculated from the number of A-Modes seen during DT divided by the total DT test time.
- Comparison of DT-I to OT-I would be used to estimate the impact of operational realism on reliability \(\Delta_{\text{DT to OT}}\).

3.1 Quantifying the failure rate for fixed failures

The dramatic reliability improvement in DT-II was the result of zero reoccurrence of fixed failure modes. Since none of the fixes eliminated the possibility of failure, we felt it was unreasonable to assume that these fixed modes would never reoccur. Instead we felt that the zero reoccurrence rates were an artifice of test duration. Applying the Gray-Lewis test for comparing two exponential means for individual fixed failure modes resulted in p-values ranging from 0.553 to 1. In other words, DT-II was not sufficiently long to conclude with any statistical confidence that the individual mode failure rates had been reduced, much less eliminated, by corrective actions.

While test data did not support conclusions about the
quality of the corrective actions individually, in aggregate was a different story. When treated together, the Gray-Lewis test resulted in a p-value of 0.0001. In other words, DT-II was long to conclude with extremely high statistical confidence that collectively failure rates had been reduced.

We decided to evaluate the fixes in aggregate by determining an aggregate Fix Effectiveness Factor (FEF) corresponding to an 80% lower confidence bound determined by a binomial distribution approach.

Using an 80% confidence bound resulted in an aggregate FEF of 0.88. Since using an aggregate FEF greater than 0.85 would have been unprecedented, we decided to set the FEF at 0.85, which corresponded to an 83% lower confidence bound. Because of the high statistical confidence in the Fix Effectiveness Factor, we decided not to run sensitivity analysis around the Fix Effectiveness Factor.

3.2 Quantify the rate of uncovering unseen failure modes.

For finite test durations of new systems, you know that if you test long enough, you will uncover an unseen failure mode. The AMPM model predicted that DT-II would uncover 3.8 new failure modes, using the data from DT-I. In DT-II, three new B-Modes were uncovered. Because of the close agreement between the AMPM prediction and the DT-II test data, we felt confident in using AMPM predictions in our projection.

3.3 Quantify the reliability degradation from DT to OT

As discussed in Section 2.1, the reliability drop from DT to OT was the result of an increased failure rate of Component A when operated by Soldiers in an operationally realistic environment. In an operationally realistic environment, the failure intensity was eight times the intensity of the more tightly controlled developmental test environment. The root cause of the Component A failures was integration geometry. The integration geometry required following specific instructions to avoid damage during normal operations. The corrective action to address this issue was to modify the instructions to reduce the likelihood of damage. Since the fix did not involve a physical change, we could not justify assuming that the fix would hold up under realistic operating conditions. The question then became what would be a reasonable Component A degradation factor for OT-II. Since there was significant uncertainty around estimating the degradation factor, we decided to do a sensitivity analysis around the degradation factor. The base case would be based on the observed point estimate DT to OT degradation of 0.88.

4 RESULTS

As we suspected, the projected reliability in the next test phase was highly dependent on the assumed Component A degradation factor. Figure 4 shows that a relatively small change in degradation factor (0.88 to 0.75) resulted in a very large change in projected reliability (MMBF = 640 miles to 860 miles). Based on historical data, we projected the likely reliability in the next test phase to be between 640 to 860 miles, falling short of the 1000 mile threshold.

Since we were projecting that we would likely fall short of the threshold, should the test be delayed? Despite our low projected reliability, we recommended continuing with the test for two reasons. First, we had identified a training opportunity that could substantially impact system reliability. Better training the Soldiers on proper handling of Component A would have a substantial impact on system reliability, boosting the system reliability above the 1000 mile threshold. Second, we felt that a second operational test was needed to validate the effectiveness of the Component A corrective action in an operational environment. It was decided to proceed with the test. We are awaiting the results so that we can compare our projections to the actual results.

REFERENCES


BIOGRAPHIES

Carol L. Vesier, PhD
U.S. Army Evaluation Center
Reliability and Maintainability Directorate
4120 Susquehanna Ave.
Aberdeen Proving Ground, MD 21005 USA
e-mail: carol.vesier@us.army.mil

Carol Vesier received her Ph.D. from Georgia Institute of Technology in Chemical Engineering in 1991. She joined the Reliability and Maintainability Directorate in November 2009. Prior to that, she was an independent consultant; assisting Fortune 500 companies improve their profitability through reliability improvements. She has developed and taught
workshops on reliability analysis techniques, as well as published numerous papers. She also developed Profit Driven Reliability™, a work process with metrics that links reliability to financial statements.

Angelo Christino  
U.S. Army Evaluation Center  
Reliability and Maintainability Directorate  
4120 Susquehanna Ave.  
Aberdeen Proving Ground, MD 21005 USA  
e-mail: angelo.christino@us.army.mil

Angelo Christino received his B.S. in Mathematics from Pennsylvania State University in 2005. He joined the Reliability and Maintainability directorate in March 2006, and since then has led the reliability evaluation efforts for major acquisition programs within the Army. He has briefed senior leaders throughout the Army and Department of Defense on system reliability planning and analysis.

Brian Hampton  
U.S. Army Evaluation Center  
Reliability and Maintainability Directorate  
4120 Susquehanna Ave.  
Aberdeen Proving Ground, MD 21005 USA  
e-mail: brian.g.hampton@us.army.mil

Brian Hampton served on active duty in the United States Marine Corps for six years before earning a B.S. in Chemical Engineering at Michigan Technological University in 2003. Prior to joining the Reliability and Maintainability Directorate in January 2010, he was a test engineer for a defense acquisition program management office, with responsibilities to include failure tracking and analysis.