Some Aerospace Engineering Applications of Reliability Growth: A Review

WALTER THOMAS, III
NASA Goddard Space Flight Center, Greenbelt, Maryland 20771, U.S.A.

(Received on October 12, 2011, revised on May 12, 2012, and July 17, 2012)

Abstract: Reliability growth analysis is a tool for trending and monitoring component and system reliability during design, development, and fielded operations. It is robust with mixed failure modes and changing reliability levels and can predict future failures with reasonable accuracy. Several reliability growth case studies from aerospace engineering applications are presented as an overview to illustrate its utility for development, in-service, and management trending and risk assessment.

Keywords: Crow-AMSAA analysis, event trending, reliability growth, risk assessment, system failure analysis

1. Introduction: The Crow-AMSAA (CA) Reliability Growth Model

In the early 1960s, J. T. Duane of General Electric derived an empirical method based on learning curves to determine when aircraft electric motors and other components under development would meet reliability goals prior to being put into production [1]. He observed cumulative failure rate vs. cumulative operating hours plotted on a log-log scale yielded linear fits for failures vs. operating time. Dr. Larry Crow improved on the Duane model by deriving a statistically based reliability growth model that provides confidence bounds, goodness of fit measures, and prediction capabilities. His works resulted in MIL-HDBK-189 [2], later updated as AMSAA-TR-652 [3]. Two International Electrotechnical Commission standards [4,5] provide cases and statistical formulae for using the Crow-AMSAA (CA) model. A U.S. Air Force study [6] considered more than twenty and studied in detail five different growth models and determined the CA model to be best overall; that study included systems having both mechanical and electronic components.

Mathematically, the model uses a power law function to relate cumulative failures to cumulative time:

\[ N(t) = \lambda t^\beta \]  

where \( N(t) \) is the cumulative number of events (i.e., failures) at time \( t \).

Taking logarithms of both sides of (1) yields:

\[ \ln N(t) = \ln \lambda + \beta \ln t \]  

where \( \beta \) is the shape parameter (slope) and \( \lambda \) the y-intercept at \( t = 1 \).

Plotting cumulative events (e.g., failures) vs. cumulative time on log-log paper should exhibit a linear fit; if it does, then the CA model is applicable\(^1\). The resulting shape parameter (\( \beta \)) indicates whether events being analyzed occur more frequently (\( \beta > 1 \)), less

---

\(^1\) Army Material Systems Analysis Activity, the organization where Dr. Crow worked when he developed CA.

\(^2\) Other methods for determining CA model applicability are described in [4,5].
frequently (β<1), or are unchanging (β~1) with time. If β=1, the CA model is equivalent to a homogenous Poisson process (HPP); if β≠1, it is a non-homogenous Poisson process (NHPP).

CA analysis models a process or system. That is, it accounts any and all intrinsic and extrinsic factors that influence events being studied. This includes the components or system being evaluated, the detection and/or measurement process involved, any maintenance actions or other changes to the system (intended or unintended!), external factors such as environment or usage condition changes, and any other factors influencing component or system reliability. It is robust and able to analyze mixed failure modes, processes or systems with reliability changing over time, and intervals with missing or suspect data [7]. It provides a visual plot depicting process or system changes along a timeline; i.e., the picture is the message. Hence, it is valuable for trending reliability and safety events to make them visible and manageable, tracking progress towards reliability goals, measuring and displaying process or system change effects, evaluating corrective action results quantitatively, and providing risk predictions. CA typically can detect changes much more quickly than, for example, moving average statistics.

When analyzing components or systems using CA, one may observe cusps (inflections), jumps3, or other changes from a linear fit on the log-log plot. These indicate change occurring in the system. This is part of the analysis. Preferred, from a reliability perspective, are cusps showing growth (shape parameters <1) rather than degradation. For processes or systems exhibiting a linear fit, future predictions of failure rates, either to the end of the test program or six to 12 succeeding months (for field data), are accurate to 10-20 percent [6,7]. Implicit for such future predictions is that no subsequent changes occur in the process or system.

A relative process or system reliability indicator is the shape parameter, β. Typical industry development processes exhibit growth slopes of 0.4 – 0.6 [8]. O’Connor [9] provides a more discerning interpretation of growth slopes, classifying them according to priority placed on reliability programs. Shape parameters for fielded components or systems have been observed from 0.1 to 4 or more.

Additional details on CA statistics, including goodness of fit, confidence interval estimates, growth or degradation determination, and future predictions are provided in international standards [4,5]. Some assert CA is applicable only to repairable systems; however, this author and others [7, 10] have demonstrated its utility for fielded non-repairable systems.

Spacecraft component and system problems (usually failures) often have large consequences and typically limited data are available. CA analysis has proven effective in trending, identifying, and tracking a variety of such engineering problems. Several examples follow. (For all graphs in this paper, italic numbers adjacent fit lines are shape parameters.)

---

3 “Jump” is used to describe a change in which a following (later time) fit line does not extrapolate back to the end point of the preceding fit line. It usually indicates a marked process or system change, assuming little or no missing data.
2. Tracking Reliability Growth during Development

Case 1: Space Meteorological Instrument Test-Analyze-and-Fix (TAF) Manufacturing Failures

A space instrument was undergoing I&T (Integration and Test) after manufacture. The contractor was performing test, analyze and fix (TAF, also called TAAF). Failures occurred in electronic card assemblies; failure modes were identified and assemblies were replaced. Cumulative failures vs. cumulative test hours were analyzed to provide feedback to the project. The resulting CA plot, Figure 1, was fit using regression methods [7].

Periods of both growth ($\beta<1$) and degradation ($\beta>1$) were observed. Between 700 and 800 hours operating time, six failures occurred causing a change from growth ($\beta=0.8$) to degradation ($\beta=2$). Management ordered a stand-down to formulate and implement remedies for those failures. After those fixes were implemented, excellent growth ($\beta=0.14$) occurred; the change to growth confirmed fix effectiveness.

The project and its customer (a federal agency) preferred the trend plot presented as failure rates. Cumulative failure rate vs. cumulative operating time is shown as Figure 2. The overall decrease in failure rate (mostly downward sloping fit lines) indicated that most fixes were effective; this was evident especially after the stand-down.

The ~2200 hour failure was caused by a series circuit card assembly capacitor short-circuiting. The assembly was replaced and testing continued. The question then was how long must the instrument be operated to demonstrate no degradation and reliability growth? As CA analysis uses only cumulative failures (successes are accounted as suspensions, or non-failed operating times for continuous data), a conservative estimate can be made by assuming imminent failure on the next succeeding

---

4 Growth vs. degradation tests were performed per [4].
time interval. Assumed future failure times were input into the software iteratively to determine at what future times $\beta$ returned to 1.0 (no degradation) and 0.6 (priority growth). Corresponding confidence interval estimates were computed about $\beta$ and times for the upper interval $\beta$ to be less than or equal to 1.0 and 0.6 determined; confidence intervals were derived using [5]. The resulting median future time to achieve no degradation ($\beta \leq 1$) was an additional 130 hours (from the failure time) and to achieve growth ($\beta \leq 0.6$) 220 hours. The corresponding times at 95% confidence were 220 and 245 hours, respectively.

This case illustrates using CA analysis to verify engineering fix effectiveness and to predict additional operating times needed to demonstrate no degradation and reliability growth.

**Case 2: Space Mechanical Component Development and Life Tests.**

A new design atmospheric sampling mechanical pump was being developed for an interplanetary discovery instrument. Two distinct requirements existed: a mission total operating time of 550 hours and 450 total pump turn-on/turn-off cycles. Development test and engineering model (EM) life test data were analyzed using CA. Since mission goals were stated as mean-times-between-failure (MTBF; hours and cycles), analysis results were converted to MTBF plots, with 95% upper confidence bounds.

MTBF for operating hours is shown in **Figure 3**. The horizontal dashed line indicates the mission requirement (550 hours) and the vertical dotted line delineates development and EM tests. The severe MTBF decrease ($\beta=49$) at approximately 4000 hours during development occurred when many different design changes were made within a few tests. The endpoint median MTBF for development tests was just below its requirement; bearing retainer and lubrication-related failure modes had predominated through the first 6000 hours of tests. Those modes were corrected. EM life tests subsequently were performed and pumps began failing by a different mode ($\beta=12$). The upper 95% confidence interval on the MTBF falls short of the requirement – by about 25 times. Even before the EM tests ($\beta = 1.2$ segment), the upper 95% MTBF was 4 times less than the requirement. The MTBF at 95% confidence not meeting the 550-hour reliability requirement indicated a high risk against attaining mission success – at the time the analysis was completed.

---

5 Dr. Robert B. Abernethy and Mr. H. Paul Barringer previously had suggested the method to the author for evaluating launch vehicle reliabilities.
The overall pattern MTBF for pump on-off cycles was similar to Figure 3. Reliability growth improved during the later development period ($\beta_s$ of 0.9 and 0.3 at 3000-7000 pump cycles) compared to degradation for pump time ($\beta = 1.2$ at 4000-7000 hours). Subsequent EM tests also exhibited degradation ($\beta = 4.9$). For pump cycles, even with the EM test failures, the median MTBF did meet its requirement. However, the 95% confidence bound MTBF was 28 times less. Trend similarity for both analyses (operating time and on-off cycles) was expected since singular tests were conducted on each tested pump. Hours- and cycles-to-failure times were extracted separately to accommodate the differing requirements. Additional engineering evaluations and tests were planned after the EM failures to address the new failure mode and provide revised pump reliability estimates.

Critical component (or subsystem) reliability evaluations in high-risk, high value scenarios must include confidence interval estimates, as shown by this case. Failure to include appropriate confidence intervals may lead to assessing requirements compliance incorrectly.

3. Identifying Problems through Chronological Performance Trending

Case 3: Space Shuttle External Tank Level Sensors System

Anomalous engine cut-off (ECO) sensor indications during fueling of Atlantis (STS-122) caused two launch delays in early December 2007. Similar anomalies during a subsequent External Tank (ET) liquid-hydrogen fueling test nine days later initiated an intensive engineering investigation to determine the cause or causes before the Shuttles would be returned safely to flight. The ECO sensors and their corresponding signals to the Shuttle’s computers are critical for a safe launch: they detect any low fuel conditions occurring prior to main engine cut-off (MECO) and immediately shut down the main engines to avoid catastrophic failure. A NASA Engineering and Safety Center (NESC) investigation after an STS-114 ET fueling anomaly identified several issues with the sensors. Corrective actions were recommended and some fixes implemented.

ET level sensor system anomalies over the program’s entirety were compiled and CA analysis was used to trend the system’s historical performance. ET-side$^7$ system operational anomalies, shown as a rate plot in Figure 4, exhibited excellent reliability growth through the first 72 flights$^8$ ($\beta = 0.27$): system anomalies were decreasing with missions flown. However, after the 90$^{th}$ flight anomalies showed an increased occurrence rate, evidenced by a inflection indicating an adverse system change. The area around the inflection point is magnified in Figure 5. This plot indicates clearly the system’s change to degraded performance. The dot-dash line from the growth fit line segment is what would have occurred with no degradation. Change from growth to degradation was detected by the visual clue prior to it being shown as statistically different: median

---

$^6$ STS is an acronym for Space Transportation System; the number designates the mission.

$^7$ Meaning those system anomalies only from the Orbiter-ET interface and within the ET were analyzed; Orbiter-side wiring and electronics had been tested and verified as not contributing to recent anomalies.

$^8$ The timeline for these analyses was mission flight number (MFN); i.e., sequentially ordered Shuttle flights. Most descriptions of Shuttle flights are referenced by “STS numbers,” e.g., STS-114, which are not necessarily ordered in the sequence missions were flown.
slopes changed to 0.87 at the 90th flight (STS-90) and 1.06 at the 101st flight (STS-97). A statistically significant change (shape parameters compared at 90% confidence) from growth was not indicated until five flights later (the 106th flight, STS-105).

This reliability growth analysis was performed retrospectively, identifying when the sensor system changed from early growth to later degradation. Change occurred more than 15 flights before the STS-114 anomaly investigations and more than 30 flights before the STS-122 anomalies. Had concurrent CA trending of system anomalies been performed, an astute engineer could have used the visible change in the plot to detect the adverse trend much sooner than when events subsequently forced a costly stand-down.

Subsequently the program inquired whether the sensor system’s fuel (H₂) and oxidizer (O₂) subsystems differed. Analyses were performed using the same data grouped by subsystem. Both subsystems exhibited similar overall patterns: growth followed by degradation. Degradation began earlier in the O₂ subsystem, at MFN 90, versus MFN 101 for the H₂ portion. Table 1 summarizes the three cases.

The 2007–08 engineering investigation identified the H₂ subsystem anomalies proximate cause to be intermittently interrupted ET electrical connector signals caused by thermal changes during pre-launch fueling. Corrective action replaced the separable connector mating contacts with permanent connections⁹.

---

⁹ Though this fix appears apparently simple, extensive testing and certification of both the revised process and personnel performing the work were required to assure no adverse impact to this safety-critical STS system.
Table 1: Summary of ET Level Sensor System Operational Anomalies

<table>
<thead>
<tr>
<th>Data Set</th>
<th>MFN</th>
<th>β</th>
<th>Significance Indication</th>
</tr>
</thead>
<tbody>
<tr>
<td>All anomalies (H₂ and O₂)</td>
<td>1-72</td>
<td>0.27</td>
<td>Reliability growth &gt;99%</td>
</tr>
<tr>
<td></td>
<td>90-121</td>
<td>1.9</td>
<td>Degradation 72%</td>
</tr>
<tr>
<td>H₂ anomalies only</td>
<td>1-72</td>
<td>0.39</td>
<td>Reliability growth 98%</td>
</tr>
<tr>
<td></td>
<td>101-121</td>
<td>3.0</td>
<td>Degradation 93%</td>
</tr>
<tr>
<td>O₂ anomalies only</td>
<td>1-72</td>
<td>0.18</td>
<td>Reliability growth &gt;99%</td>
</tr>
<tr>
<td></td>
<td>90-121</td>
<td>2.5</td>
<td>Degradation 62%</td>
</tr>
</tbody>
</table>

STS-122 was launched successfully in February 2008 with no ECO anomalies, as were two preceding launches. The number of future missions flown failure-free needed to demonstrate no degradation and reliability growth in the system were determined using methods outlined in Case 1, above. For All (H₂ and O₂ subsystems inclusive), no degradation would occur four launches after fix implementation (one additional launch after the three successive anomaly-free ones) and growth after nine successive launches (five additional). No degradation would be demonstrated in the H₂–only portion after six post-fix launches (three additional) and growth after twelve post-fix launches (nine additional). The H₂–only part of the system required more flights to demonstrate no degradation and reliability growth because its occurrence rate during degradation was more severe than for All (the H₂–only slope, β, was 3.0 vs. 1.9 for All). This case shows that tracking safety-critical events proactively (vs. retrospectively) can identify potential problems early, allowing engineering analysis and fixes before problems become significant. Total system behavior (i.e., All: both H₂ and O₂ subsystems) also was shown to be divisible into sub-elements to determine if system elements or components contribute differently to total system behavior.

Case 4: Space Shuttle Orbiter Lower Surface Damage Trends

After each Orbiter flight the vehicles are inspected for thermal protection tile damage; damage is measured and counted by size greater than or less than or equal to 1.0 inch. Cumulative tile damage counts per mission for all vehicles (pooled) exhibited both growth and degradation; shape parameters (β) ranged from 0.6 to 3.8, Figure 6. An engineering goal would be correlating observed damage rate changes with system changes, for example, launch environmental conditions, mission profiles, orbital exposure (micrometeoroids, other space debris), vehicle dynamics, reentry conditions, ET and vehicle pre-flight processing. The early slope change (from β=1 to 3.8) corresponded with replacing the original tank with a lightweight design. System interaction complexities, however, precluded identifying and correlating causes for most of the other observed changes. Damage counts (lower surface damage >1 inch) then were differentiated by vehicle and analyzed using mission flight number (MFN) as the time variable. Results for Discovery (OV103) and Endeavour (OV105) are shown in Figure 7. Damage accumulation rates for OV103 and OV105 differed significantly. OV103 incurred mostly
decreasing damage rates ($\beta=0.5, 0.22, 0.55$) after initial degradation ($\beta=4.4$). OV105, however, exhibited increasing damage rates ($\beta \geq 2.4$) over its lifetime. Damage rate increases for both vehicles around MFN 90 corresponded with an ET insulation process change. OV104 (Atlantis) exhibited a pattern nearly identical to OV103’s except no similar increase around MFN 90 was seen; it was undergoing major maintenance and did not fly when the ET process change occurred.

This case illustrates that total system behavior may be confounded if event data for multiple units are combined and that analyzing by individual units can determine if they behave differently, as in this case, or similarly, as in case 3.

---

**Case 5: Trending Deficient Field Data to Estimate Supplier Quality**

During the Shuttle ECO investigation (Case 3, above), attention focused on the connector interfacing ET sensor signals with Orbiter electronics systems. Engineers desired to know if the connector manufacturer’s outgoing product quality was suspect when parts were supplied to the program. Ideally, one would review manufacturing yield and/or customer return data to estimate product quality. However, those data are proprietary and typically unavailable. Outgoing quality can be assessed qualitatively by trending GIDEP problem reports, advisories, and alerts associated with a given manufacturer (or manufacturers), under an assumption that GIDEP report frequency reflects outgoing quality. (“Deficient,” for this case, means available data were only the number of GIDEP reports; fielded product population sizes were unknown.)

---

10 GIDEP (Government-Industry Data Exchange Program) is a cooperative activity between US government and industry participants seeking to reduce or eliminate expenditures of resources by sharing technical information essential during research, design, development, production and operational phases of the life cycle of systems, facilities and equipment. Various failure and other data are collected and disseminated for use among participating organizations and companies. See www.gidep.org/.
GIDEP reports assignable to the manufacturer were summed cumulatively by year, with results in Figure 8 as a CA rate plot. Reports exhibited a constant rate during 1970-1979. The number of reports increased significantly ($\beta=3$) in 1979-81 suggesting quality had deteriorated. Nominal growth ($\beta=0.8$) occurred for the next seven years. After another period of degradation, no GIDEP reports were recorded during 1992–97 and excellent growth occurred thereafter ($\beta=0.24$). It was concluded that likely there were no general quality issues with the manufacturer during 1995–2003, when the program procured its connectors.

Is the assumption that GIDEP reports reflect manufacturing output quality valid? A NESC team subsequently investigated electrical connector applications and processes across the Agency [11]. Connector problem reports were compiled from the GIDEP system and from across four NASA centers. CA analysis compared GIDEP vs. NASA problem reports, Figure 9. Overall GIDEP and NASA trends were identical. (Shape parameters were not significantly different at 90% confidence.) Thus, for electrical connectors, GIDEP reports did represent a sample of connector problems among the four centers (a population). GIDEP reports were, however, under-represented by about 70%.

This technique has been used for numerous aerospace part, component, and system problem assessments. One benefit over, e.g., numbers in a spreadsheet or typical line or bar charts is growth or deterioration is indicated clearly by the CA shape parameters. Growth or degradation significance also can be tested statistically. However, occurrence probabilities cannot be extracted because fielded population sizes typically are unknown.

Figure 8: Connector Manufacturer GIDEP Reports per Annum.

Figure 9: Connector Manufacturer Problem Reports from GIDEP and NASA compared.

11 Only those records that would warrant a GIDEP submission were included for this comparison; these were problems or failures attributable to connector manufacturers.
3. Event Trending for Management Visibility

Case 6: Safety Event Occurrences for Management Monitoring

At one aerospace facility, safety events (close calls, incidents, and mishaps) associated with space flight hardware manufacturing and testing increased significantly. A spreadsheet line chart showed several apparent changes in the occurrence rate of events per month over a four-year period. Typically, line and bar charts generated to describe such trends, however, tend to focus attention to event (occurrence) magnitudes instead of frequencies. A CA rate plot, Figure 10 more clearly showed event trend occurrences over time. From January 2004 to March 2005, safety events per month decreased ($\beta=0.8$), followed by a significant increase ($\beta=2.6$) over the next three months. The event rate subsequently decreased ($\beta=0.4$) until another increase began in November 2007. Increased facility loading (more systems concurrently tested) caused the later adverse change. Cause of the March-June 2005 increase was not determined (labor hour and staffing data were not available for that period).

This case illustrates that CA analysis can provide clarity in assessing events significant to management. Concurrent trending can measure system behavior quantitatively and track management action and change effects.

4. Conclusions

Crow-AMSAA (CA) analysis applications involving part, component, and system issues encompassing design, development, and operations in aerospace applications have been presented to demonstrate broad applicability of reliability growth methods. Uses included verifying engineering fix effectiveness, predicting future test and operating times to demonstrate growth, assessing reliability requirements compliance, determining component or subsystem contributions to total-system reliability, assessing production output quality when detailed manufacturing data are unavailable, and clarifying safety event trends for management visibility. CA trending detected adverse system changes before existent failure monitoring methods, indicating concurrent trending can provide early warning of impending system problems. CA analysis can provide quantitative component and system performance assessments, enabling more accurate and conclusive risk and reliability evaluations.

Acknowledgements: NASA colleagues contributing to these studies include: Mr. Robert Beil (NASA Engineering and Safety Center); Mr. Mitchell Davis, Mr. Roger Farley, Mr. Stephen Glockner (SGT Inc.), Mr. Robert Kichak (Muniz Engr.), Dr. Henning Leidecker, Mr. Ed Rutkowski (SRS Technologies), Mr. Michael Sampson, and Ms. Mary Walker (Goddard Space Flight Center); Mr. Ray Gomez and Mr. Vince Reyes (Johnson Space Center); Dr. Chris Iannello and Hung T. Nguyen (Kennedy Space Center). Dr. Robert B. Abernethy introduced the author to CA analysis; he
and Mr. Paul Barringer (Barringer and Assoc.) have encouraged, advised, and critiqued these and other risk and reliability applications. Mr. Tim Adams (Kennedy Space Center) and Mr. Ken Johnson (NASA Engineering and Safety Center) reviewed the manuscript and provided illuminating suggestions.

References


Walter Thomas, III received the B. Ceramic Engineering from the Georgia Institute of Technology (1969) and M.S., Ceramic Engineering, from the University of Illinois (1972). Prior to joining NASA in 1986 he worked in glass and electronic materials, component, and packaging industries. At NASA’s Goddard Space Flight Center he has held engineering and supervisory positions in electronic parts, packaging and process engineering, and reliability and risk assessment, providing assessments for the Hubble Space Telescope, Space Shuttle, International Space Station, and Constellation programs, and numerous other national asset satellite programs. His interests include life data, reliability growth and other statistical analyses, field performance evaluations, complex systems modeling, and technical glass applications and failure analysis. He presently is a Senior Reliability Engineer at the Goddard Space Flight Center and a member of the American Society for Quality.