IMPROVEMENTS IN AUTOMATED RELIABILITY GROWTH PLOTTING AND ESTIMATIONS

David Dwyer*, Edward Wolfe, and Jonathan Cahill§**
BAE Systems, Inc., Nashua, NH

Key Words: *Duane, MTBFc, MTBFi, Reliability Growth, Test-Analyze-Fix*

SUMMARY & CONCLUSIONS

Calculations of Instantaneous MTBF have gotten sloppy since J. T. Duane [1] and E. O. Codier [2] published their papers on the subject 43 years ago. Codier made the following notes on plotting the line through the points in his definitive description of Duane's Reliability Growth calculation methods that he presented at the 1968 Annual Symposium on Reliability:

- The latter points, having more information content, must be given more weight than earlier points and
- The normal curve fitting procedure of drawing the line through the "center of gravity" of all the points should not be used.
- Unless the data is exceptionally noisy, the best procedure is to start the line on the last data point and seek the region of highest density of points to the left of it.

These principals for estimating MTBF from non-homogeneous data are not being followed and the result is a less than accurate estimation of current (aka, instantaneous) MTBF.

This paper describes a spreadsheet method for following Duane and Codier's original recommend-dations that will avoid the errors in judgment that are common when "eyeballing" the growth line or the greater errors that result from using the all-to-available "trend line" method. The proposed method automatically weighs the cumulative points (each successive point weighing more), calculates their center of gravity, draws a line from the center-of-gravity point through the last point, and calculates an instantaneous MTBF. Calculations have been made using these methods as well as the common method of drawing a line through those same points using a spreadsheet trend line (aka, least squares fit) and been applied to a multi-year field reliability program. The calculation of instantaneous MTBF (MTBFi) for each of these methods was compared with a moving average centered about successive points over the life of the program and as a control measure of MTBF.

We will show that if we adhere to the recommendations of the original papers by Duane and Codier and develop a method to weigh the latter points more, and draw the line through those weighted points and the last point, that significantly better estimates of instantaneous MTBF result. The error in calculation of MTBFi using an automated version of Duane and Codier's recommendations was reduced by 34% compared with a non-weighted trend-line calculation.

Reprinted from the *Proceeding of the RAMS-2009 Annual Reliability and Maintainability Symposium, Fort Worth, TX, 26-29 January 2009, pp. 377-382*.
ISSN 1559-9388 (online version), ISSN 1559-9396 (CD-ROM version).

1 INTRODUCTION

1.1 Background

The data for this paper was chosen from a large development program during initial field testing where there was a significant interest in reliability. It was important to decide what kind of information or data we had, given the nature of multiple products deployed with varying degrees of corrective actions implemented. Before we can attempt to plot Reliability Growth, we must decide what kind of data we are going to deal with. We will then briefly review Duane's recommendations and three approaches to drawing a trend line through the data. We will then compare the accuracy of the estimates over a three year test period using both the common least squares fit line and an automated version following Duane's directives.

1.2 Ideal Reliability Growth Data Compared With Typical Field Reliability Growth Data

We will be dealing with field data for a product that is still under development, and part of a Test-Analyze-Fix program. The probabilities of occurrence of failure modes for ideal data are distributed evenly when plotted on a log scale, like white noise. This can be seen in the observations by J. T. Duane [1] in his initial paper where he noticed that reliability growth data plotted like manufacturing learning curves. Failure modes that plot like this are independent from each other and the data non-homogeneous. The less probable failure modes in the future are completely independent from those we have seen so far. Duane showed that we can proceed in correcting these failure modes as we see them and calculate cumulative reliability as we continue in our test and fix program. He also said that the cumulative MTBF vs. test times tend to form a straight line when the data are plotted on a log/log scale. This is the ideal case where failures are corrected on the test subjects when they are discovered, not a situation typically found in field data. However, Duane's recommendations provide a method to track data that is somewhat noisy, but not so much that a new plot needs to be initiated.

1.3 Ideal Reliability Growth Data

We can generate ideal data to exercise our hypotheses on a subset of learning curve "data sets". This data will be scripted for our purposes in that the "Probabilities of Occurrence" will be distributed perfectly evenly when plotted on a log-log scale (i.e., Log MTBF vs. Log test hours). An example of such data is summarized in Table 1 and plotted in Figure 1. This will never happen in reality, but it is consistent with the "white noise" premise and it provides a data set that allows us to test some of our ideas.

Table 1: Test Times for White Noise

Failure Mode	Test Time (hours) @ failure
1	20
2	200
3	2,000
4	20,000
5	200,000
6	2,000,000

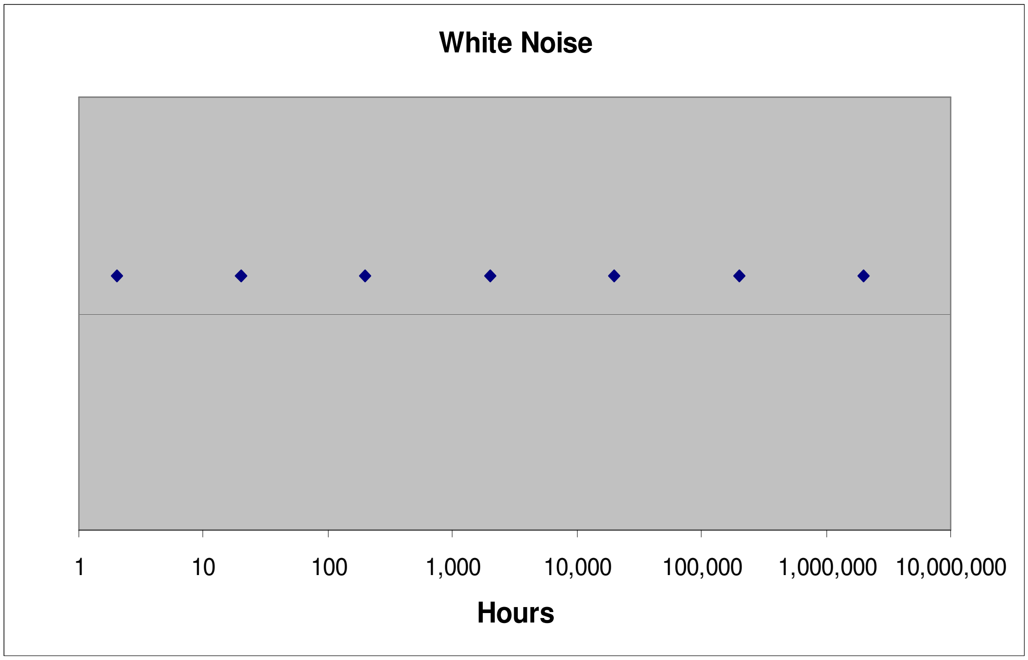

Figure 1. White Noise Failures vs. Test Hours

White noise consists of an infinite number of failure modes with the probability of each one a common multiple (on a log scale) of the previous one, +/- some uncertainty factor. When electrical white noise is displayed on an oscilloscope, successive sweeps are always different from each other, yet the waveform always "looks" the same. Each of these failure modes is completely independent from the others. The Test and Fix program is most ideal when applied to a single unit that has perfect corrective action implemented for each failure as it is discovered. But we have multiple units deployed in the field, some with process, component or design corrective actions implemented, others not. For our ideal example, the uncertainty will be zero in order to see what the effects are for the way we test, do failure analysis, etc. This will serve as a basis for some manipulations.

1.4 Effects on Growth Slope Due to Corrective Action Effectiveness for Ideal

Given a set of ideal data, let us see what the effect of two levels of corrective action effectiveness would be. Assume that we may either correct each failure as we encounter it the first time, or we correct it after we have seen each failure twice. When we need to see it twice, we will assume that twice the amount of the test time for each failure mode passes before it is corrected. Then that failure mode will never occur again. Both instances are summarized in Table 2 and plotted in Figure 3. The surprising result is that correcting each failure the first time we see it (aka, 100% corrective action rate) does not result in a different "growth rate" than correcting failures after seeing each one twice (50% corrective action rate). It just displaces the curve to the right. Of course, the displacement is across a log scale. The reason it seems surprising is that we are not used to thinking in logarithmic terms.

Table 2: Corrective Action Effectiveness

Corrective Action Effective-ness	Test Time @ Failure	Cumu-lative MTBF	Cumu-lative Failures	Log Test Times	Log MTBF
100%	20	20	1	1.30	1.30
100%	200	100	2	2.30	2.00
100%	2,000	667	3	3.30	2.82
50%	20	20	1	1.30	1.30
50%	40	20	2	1.60	1.30
50%	200	67	3	2.30	1.82
50%	400	100	4	2.60	2.00
50%	2,000	400	5	3.30	2.60
50%	4,000	667	6	3.60	2.82

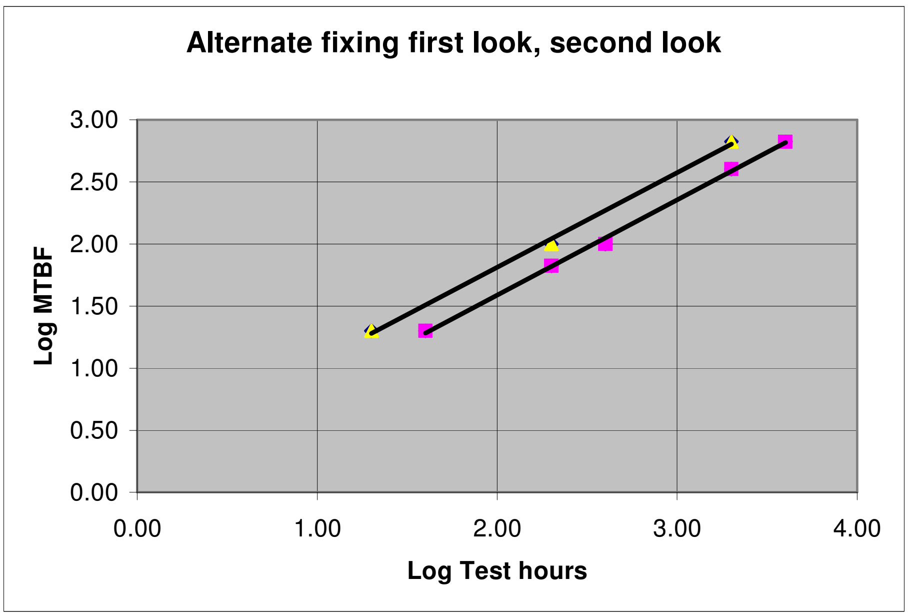

Figure 2. MTBF Growth vs. Corrective Actions

1.5 Real World Data

Unlike our "ideal" where failure mode probabilities are spaced evenly on a log scale, typical field failure modes will have significant variability in spacing of probabilities. These are independent events so removing failure mode No. 1 from the design does not remove failure mode No. 2, etc. Current reliability growth estimates will ultimately be verified by moving average estimates that have not yet been seen. We know from experience that the spacing of these failure modes on a logarithmic scale is 'roughly' even. However, there are several sources of variability in the spacing of these failure modes along the log axis besides the natural differences in the probabilities of the failure modes themselves. These include:

- Corrective actions are not perfect so failure modes are not removed uniformly.
- Simultaneous testing of multiple units in the field adds variability.
- Not all of the fielded units have the same corrective actions implemented.

- Testing of fielded items adds other variances:
- Time to determine corrective actions for field returns may be long compared to the MTBF.
- Design changes can significantly affect reliability of new items put in the field along side existing ones (could be better or worse).
- Field units added regularly give variability of age to the test sample.

2 RELIABILITY GROWTH MODEL DEVELOPMENT

2.1 Our Requirements for Handling Noisy Field Data

Since variances cause irregularities in the data plots, we will consider it part of our requirement to deal with these sources of variation. We will show an example wherein all of the sources of variation exist but that the calculated estimate we use adapts to and that it is far superior to using a least squares fit. We will use data from a field reliability growth program and plot Reliability growth based on the learning curve methods proposed by Duane and compare the results with estimates using a least squares fit. Our reliability growth model requirements are:

- The growth model must allow us to estimate current reliability with all the variability inherent in our test environment.
- Root cause and corrective action will be determined for each failure but:
- Not always the first time and
- There will not always be a recall when corrective action is identified.
- We will plot MTBF against total system accumulated operating hours.
- With each ‘fix’, the residual sample of failure modes and their probabilities is changed, so the test sample will not be homogeneous.
- Judgment factors need to be removed from the estimating process, (such as ‘eyeballing’ the data to determine where a line should be drawn) so the calculations will have to be done automatically.

Because of:

- variability in the data and
- because the test population is not homogeneous and
- the estimate is for reliability at some future time when new failure modes will determine reliability, we will not calculate confidence on these MTBF estimates.

2.2 Duane’s Recommendations

J. T. Duane provided the following example: “In an effort to determine the manner in which reliability performance changes during development and design improvement activity, data was analyzed for a total of five different products. Available reliability data was analyzed in search of consistent patterns which might apply for a wide range of equipment types. A remarkably consistent pattern finally emerged when cumulative failure rate (defined as total malfunctions since program start, divided by total operating hours since start) was plotted on log-log paper as a function of cumulative operating hours. Figure 3 from Duane’s paper shows the change in cumulative failure rate over operating hours, plotted on a log-log graph.

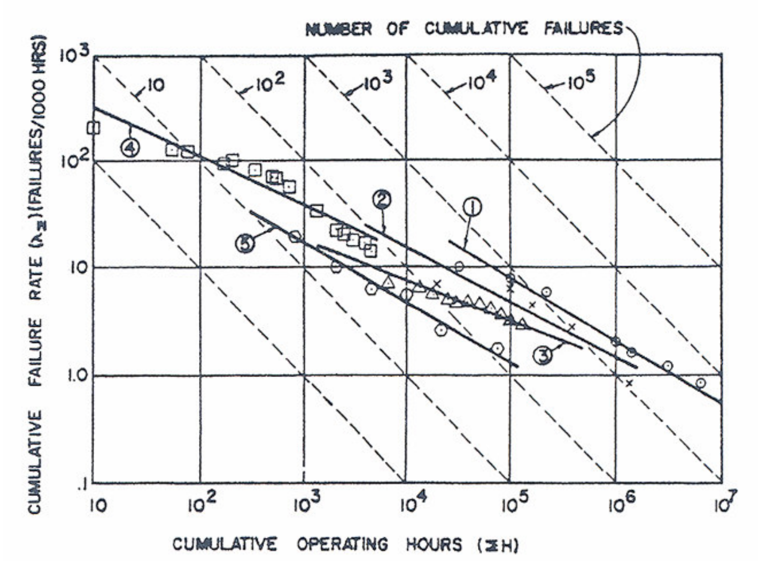

Figure 3. Planned Improvements in Reliability

The end point includes all of the information up to that time and because of this, we can have the best calculation of cumulative failure rate there. Our Method must be able to calculate current reliability based on the cumulative reliability at the last point and the fact that the data tends to fall on a line. These plotting criteria were described by E. O. Codier, [2] a contemporary of Duane at General Electric Aerospace, in his paper presented at the 1968 Annual Symposium on Reliability:

- "The latter points, having more information content, must be given more weight than earlier points and
- The normal curve fitting procedure of drawing the line through the 'center of gravity' of all the points should not be used.
- Unless the data are exceptionally noisy, start the line on the last data point and seek the region of highest density of points to the left of it."

He also said that "...for presentation purposes..." we should plot a point which includes all the test time accumulated up to "time now", even though the last failure occurred some time ago. He added that it "...yields a slightly optimistic point which does not have the same information content as a failure point, and should not be included in curves for slope determination."

Duane observed that the points will eventually fall on a straight line, with a slope 'm', much the same as learning curves do in time studies for tasks in a manufacturing environment. He did not say that this line will extend forever, just that for several cycles on the log-log graph, they will fall on a straight line.

There were three parameters of interest, (λ_c, ΣF and ΣH):

- Cumulative failure rate λ_c or alternatively, its reciprocal, MTBFc, cumulative MTBF. The measure for cumulative failure rate is Σ(Failures)/Σ(Hours), and is best represented by the last point on the graph, since it includes the most and latest (most 'fixes' incorporated) data. Symbolically, this is: $\lambda_c = \Sigma F/\Sigma H$
- Cumulative failure rate will vary in a manner directly proportional to some negative power of cumulative operating hours = $kT^{(-m)}$

- Instantaneous failure rate, λ_i, is the time derivative of ‘F’. MTBFi (Instantaneous MTBF) is it’s reciprocal. The full derivation for λ_i is:

$$
\begin{aligned}
\lambda_c &= F / T \\
&= kT^{(-m)} \\
F &= kT^{(1-m)} \\
\lambda_i &= \partial F / \partial T \\
&= k(1-m)T^{(-m)} \\
\lambda_i &= (1-m)\lambda c \\
m &= slope
\end{aligned}
$$

The slope of the line (on a log/log scale) was in the range of -0.4 to -0.5 as can be seen in Figure 3. J. T. Duane published a few papers around 1964 that described a method for estimating current and instantaneous reliability. His observations were simple [1]. If you are involved in a program to test hardware and find root cause and correct failures, then:

- Collect data:
- Failure count
- Hours of test time or
- Number of test samples for one-shot items, (e.g., rockets).
- Plot on a log-log scale:
- Failure count
- Hours of test time or
- Number of test samples.

As we stated initially, we must draw the growth line through the last point and through the center of gravity. We will automate the method to avoid the impulse of the casual observer to influence the calculation of current reliability by ‘tilting’ the line in his favor by changing the slope ‘m’, a factor in the reliability calculation.

These criteria, along with the equations above, constitute our method to satisfy our requirements. We will literally find the ‘center of gravity’ by assigning ‘weight’ (=’n’ for the n^{th} point) to the points (except for the last one), each successive point having more weight than the one before it. We will then draw a line through the center of gravity of those weighted points and through the last point. When we have done that, we will have complied with Codier and Duane’s guidelines and we will have satisfied our requirement to “Draw the growth line as objectively as possible”.

2.3 Least Squares Fit and the Effects of Weighing the Latter Points More

We want to be able to implement the criteria in the System Design, above, into a “Detailed Design”. Requirements 1 through 3 and requirement 5.b. in System Design talk about the latter points having more information and about the center of gravity of the points. This is a modification to the common process of making a least-square fit through all of the points. For example, the normal curve fitting procedure of drawing the line through the “center of gravity” of all the points should not be used. We will want to compare results of least squares fit to our method to be sure that there is an improvement. Dhillon [3] describes one way to do a weighted fit of the points, although he does not go through the last point. Dhillon describes both the least squares fit (aka, “trend line”) and the “weighted” least squares fit: [3, p.150].

"If the plotted points are not independent, then proportional weighting the cumulative number of failures at each point is a reasonable way to improve accuracy of these estimates. This technique assigns greater weight to the preceding data point (the most recent one). This method is based on the assumption that each data point is plotted *m* number of times at that point."

2.4 Weighted Least Squares Fit Through the Last Point

We want to go one step further than weighing the latter points more. We also want to go through the last point. We will do this by giving each point a weight according to its failure number (F), except for the last point and find the resulting "center of gravity" of those points. We will then want the line to go through the "center of gravity" and also the last point. We will then have an objective way of adhering to Duane's "notes on plotting the line through the points" without "eyeballing" the line. We will, in fact, find a center of gravity as if the points were rocks on a [weightless] board. See Table 3 for a simple example.

Table 3: Examples Data

ΣF = Wt.	ΣH	MTBFc = ΣH/ ΣF	Log (ΣH)	Log (MTB Fc)	Weight x LogΣH	Weight x Log (MTBFc)
1	25	25	1.40	1.40	1.40	1.40
2	55	27.5	1.74	1.44	3.48	2.88
3	95	31.7	1.98	1.50	5.93	4.50
4	140	35	2.15	1.54	8.58	6.18
5	200	40	2.30	1.60	-	-
10			1.94	1.50	19.40	14.96

The first four data points in the table are used to calculate the center of gravity. That point (CGx = 19.4/10 = 1.94, CGy = 14.96/10 = 1.50 for the first four points) and the last point (2.30, 1.60) will be used to draw the line and to calculate MTBFi. The resultant plot is shown in Figure 4. It is evident that a trend line would have a different slope and end point and result in a different MTBFi, but our line follows the most recent data better.

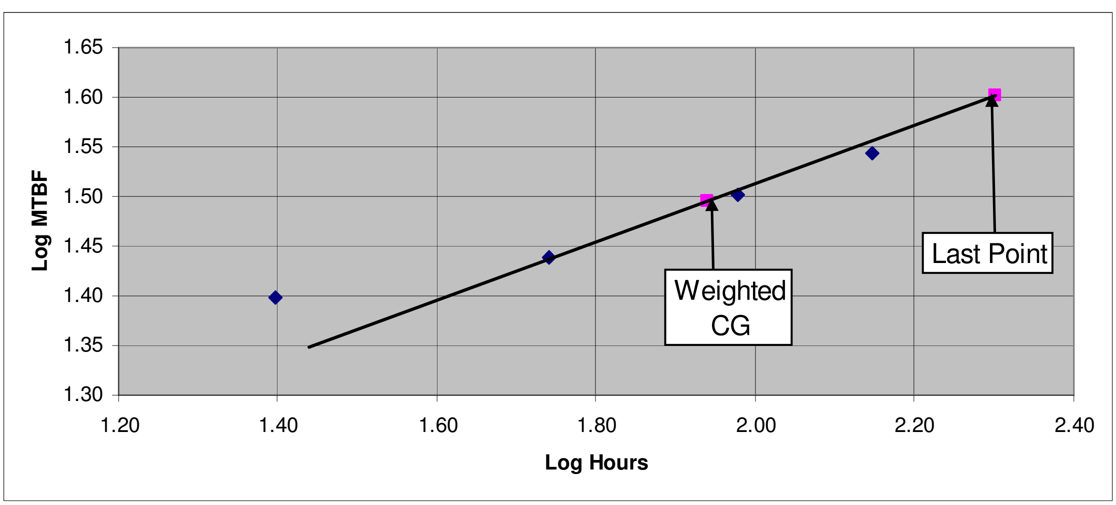

Figure 4. Sample Plot

3 STUDY RESULTS

As we stated initially, we want to be able to deal with conditions where there are many sources of variability in spacing of failure mode probabilities. In fact, we want to see if we can estimate current reliability when we have all of the following sources of variability present:

- Imperfect corrective action effectiveness
- Significant delays in implementing corrective actions into fielded units
- Component failures
- Process failures
- Design failures
- Multiple units being tested, some old, some new
- Significant design changes made to some of the fielded units but not others.
- Delayed measure of continuous improvement initiatives

These conditions have, in fact, all been present in the testing of units for the following Reliability Growth Plot, Figure 5. By emphasizing the most recent data points more and by drawing the line through the last point, Duane's methods adapt to changes on slope much better than a trend line that relies as much on old data as on new. The last MTBFi estimate has a much closer fit to the six month moving average.

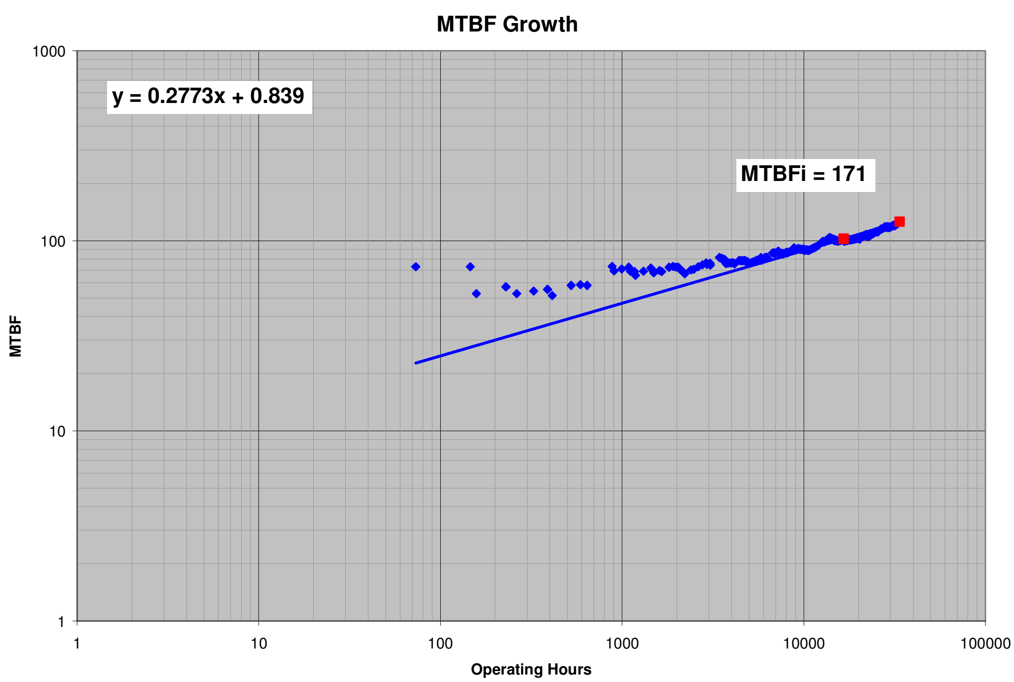

Figure 5 - Reliability Growth Plot of Fielded Units

In comparison, the least squares fit methods used against the same data do not keep up with the changes in growth slope (See Figure 6).

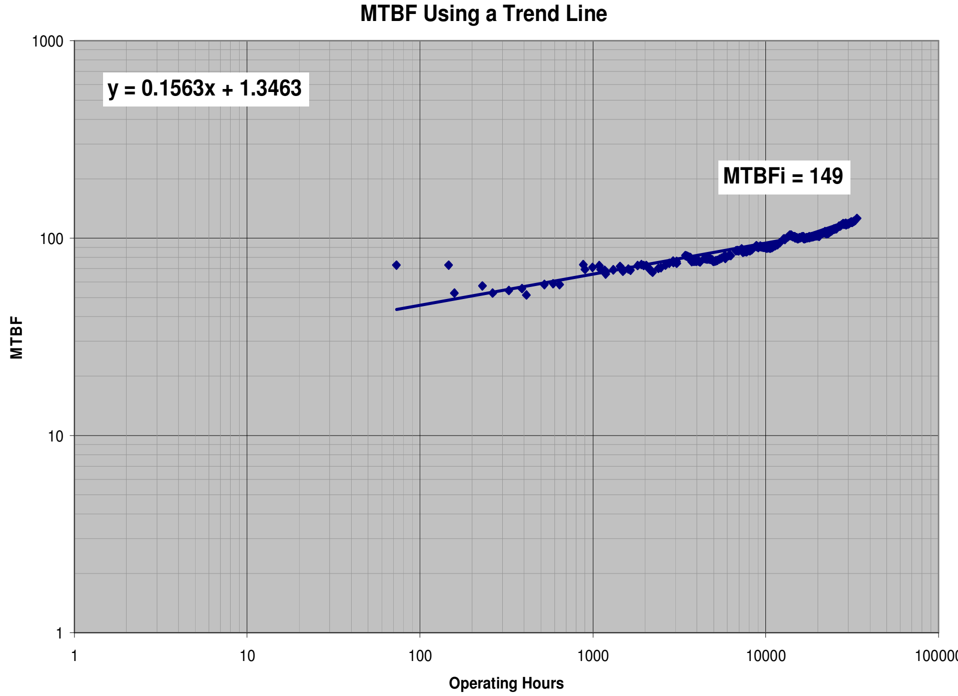

Figure 6 - Reliability Growth Plot of Fielded Units — Least Squares Fit

The effectiveness of the two methods in question can be compared with a moving average +/- three months about that point. The results and comparisons are shown in Table 4.

Table 4: Comparison of Results — Duane vs. Trend Line

Time (Months)	Moving Ave. MTBF	MTBFi, Wt'd, Last Point	Wt'd, Last Point Error	MTBFi Trend Line	Trend Line Error
0	69	118	49	69	0
3	81	69	12	78	3
6	107	100	7	81	26
9	112	103	9	86	26
12	131	137	6	95	36
15	134	131	3	102	32
18	128	117	11	112	16
21	121	129	8	117	4
24	112	124	12	119	7
27	137	139	2	128	9
30	167	157	10	138	29
33	197	171	26	149	48
Sum of errors			155		236
Average error			13		20

It can be seen from the results shown in Table 4, that the trend line estimate resulted in a 54% increase in average error compared with the weighted, last point criteria of Duane. That's because weighing the latter points more and going through the last point helps it adjust to changes in slope due to the sources of variability inherent in fielded systems still under development. The final MTBFi estimate of 171 hours compared with the 197 hour MTBF moving average over 6 months is much closer than the 149 hour MTBFi estimate using a trend line. This was for noisy unfiltered field return data where there was no objective way to decide what to censor or where to restart the plot. Different people would have chosen different failures to censor or places to restart, and that is not an objective way to estimate MTBFi. It was not always obvious what the causes in variation were or when things were stable. But that is the point – this is non-homogeneous data and everything is in a constant state of change.

Our experience is that the method initially described by Duane and Codier works best for estimating current MTBF from field development testing, especially when there are numerous sources of noise in the data. Some will say that this "noisy" data does not lend itself to MTBFi estimates, but the owner of these units wants to know the MTBF of what he has in the field and what progress is being made. We maintain that these situations do not warrant confidence calculations. It's just the best estimate available, like in learning curves for touch labor in manufacturing or cost projections for the n^{th} unit. This paper shows that for the typically noisy field test data, significant error is introduced when the conventional approach of using a least squares fit is employed and that following Duane's original recommendations for line drawing is the best way.

REFERENCES

1. Duane, J. T. Learning Curve Approach to Reliability Monitoring, *IEEE Transactions on Aerospace*, Vol. 2, No. 2, April 1964
2. Codier, Ernest O. Reliability Growth in Real Life. In *Proceedings of the 1968 Annual Symposium on Reliability*, New York, IEEE, Jan. 1968, pp. 458-469.
3. Dhillon, Balbir S. *Reliability Engineering in Systems Design and Operation*. Van Nostrand Company Inc., 1983.

---

* **DAVID DWYER** is a Principal Software Engineer and has worked both as a reliability engineer and as a software engineer while at BAE SYSTEMS over the last 37 years. Before working at BAE SYSTEMS, he was a flight instructor at the University of Illinois for 3½ years. He has a B.S. in Physics (Providence College, 1963), an M.S. in Electrical Engineering (Northeastern University, 1980), and M.S. in Computer Science (Rivier College, 1999). He has presented several papers on Software and Hardware Reliability including "Software Reliability Estimations/Projections, Cumulative & Instantaneous" (RAMS-2004 Symposium), "The Papers of Downs and Tractenberg", (58th NEQC Conference, 2010), "Improvements in Estimating Software Reliability" accepted for presenting at the RAMS-2011 Symposium, as well as "Reliability Test Planning for One Shot Systems" (RAMS- 1987 Symposium), "Hardware Reliability Growth Estimations and Projections – What is Valid and What is Not", (ASQ NEQC 56th Conference, 2006), and "Sweaty Days = Shorter Runways" published in *Private Pilot Magazine* (June 1972). He is also an adjunct Computer Science faculty member at Rivier College, Nashua NH, where he teaches graduate and undergraduate courses in software reliability.

** **EDWARD WOLFE** is a Principal Reliability Engineer and has worked at BAE SYSTEMS for the last 6 years. While at BAE SYSTYEMS he has done work plotting and estimating reliability growth for various programs from Manufacturing Development through Production phases. Before BAE SYSTEMS, he has experience working for commercial companies in Engineering, Reliability, and Management roles. He has a B.S. in Mechanical Engineering (University of Massachusetts Lowell, 1999).

§ **JONATHAN CAHILL** is a Senior RMS engineer at BAE SYSTEMS. While at BAE SYSTEMS, he has done work plotting and estimating reliability growth for various programs from Manufacturing Development through Production phases. He has a B.S. in Electrical and Computer Engineering (Worcester Polytechnic Institute, 2003).