Contemporaneous Failure Time Analysis Using Poisson Probability

Jan Smith, PE, Zachry Industrial Inc.

Oliver Sac, BSME, Chevron Phillips Chemical Company LP

Kevin Bordelon, PE, BSChE, Zachry Industrial Inc.

Key Words: Poisson, Probability, Repairable Systems, Reliability, Homogeneous Poisson Process, Non-Homogeneous Poisson Process, Monte Carlo

SUMMARY & CONCLUSIONS

In some operating process plants, equipment systems are monitored and all failure times are immediately analyzed to reveal possible onset of reliability degradation. This allows intervention and avoidance of future failures that otherwise will be experienced. Uncommon methods of failure time analysis and data visualization are required to detect the immediate onset of trends. Also, the large number of analyses are made practical only with some degree of automation.

A null hypothesis concept allows the Poisson distribution to be used in reverse to identify failures that do not fit the homogeneous Poisson process (HPP). Low probability values (p-values) for the last failure, last two failures, etc., provide evidence that special cause failures are occurring and intervention is appropriate. With the various p-values regarded as random variables, probability distributions are formed for each via Monte Carlo simulation.

Poisson analysis is aided by two data visualization methods. A probability map displays all p-values for the time-between-failure (TBF) data set. Simple conditional formatting in an Excel spreadsheet highlights any developing trends. A TBF residuals graph (abscissa being actual failure number minus failure number predicted by the data set mean and ordinate being cumulative time) sensitively displays trends and supports efficient calculation of the various p-values.

The failure time residuals graph produces a largest positive or smallest negative residual value. This transition point is a change in trend direction. It exist even if the trend is only random variation in a HPP. This transition is a data set statistic. With Monte Carlo simulation of numerous HPP data sets, a probability distribution for this statistic is formed and serves as a null hypothesis against which the data set statistic is compared. The p-value for the data set maximum/minimum residual statistic is used as strength of evidence of a data set trend.

Crow-AMSAA, Poisson, and maximum/minimum residual methods are compared using a well know data set widely cited in the reliability literature.

While most work to date has been directed at reliability degradation detection, the methods are thought to be of potential value in reliability growth.

Additional work that is needed is to describe the residual maximum/minimum statistic mathematically and to evaluate Poisson and maximum/minimum residual methods against others over a complete range of computer simulated data sets. The Poisson is obviously limited to those situations where the HPP is a valid null hypothesis.

1 INTRODUCTION

Repairable systems in the process industries can experience abrupt and unexpected increase in failure rates that often are belatedly recognized. In order to recognize a change in reliability as quickly as possible for intervention, two non-traditional methods are introduced. These are Poisson probability values (p-values) and time-between-failure (TBF) maximum/minimum residual. Both are treated as random variables with probability distributions around the mean developed with Monte Carlo simulation. These are effective with very small data sets and are efficient. Efficiency is important because to detect failure trends at the earliest possible time, analysis is contemporaneous with the failure. With thousands of assets in the typical process plant, the analysis volume is high; therefore, practicality requires the methodology be sufficiently automated. Fundamental to both methods is the use of the HPP as a null hypothesis.

A particular data set is used throughout the paper to demonstrate how p-values are obtained and how probability distributions around the p-values are obtained with Monte Carlo simulation. Multiple p-values for various look back periods to identify the significance of step changes in failure rate are introduced.

TBF residual trends and probability maps are introduced and described using the same data set. These allow very efficient analysis and visualization of the results.

A TBF maximum/minimum residual value is introduced. This value is a data set statistic, although not mathematically defined as far as we know. The null hypothesis probability distribution for this statistic is developed by obtaining numerous statistic samples from computer generated data sets known to be independent and identically distributed exponential.

Crow-AMSAA, Poisson, and maximum/minimum residual methods are compared. The comparison will emulate contemporaneous analysis where each failure uses only the then current history and is blind to the future.

2 POISSON DISTRIBUTION

The Poisson distribution is used to determine the probability of specific numbers of events occurring within a specified time interval, when the events are generated by a HPP. Failures times are independent and identically distributed exponential random variables. The mean number of events must be constant for any time interval of equal length. Repairable system failures are, in general, such a HPP. But new failure modes, improper repair, and any other special cause produces TBF data that does not fit the HPP conditions for Poisson. Moreover, it is these nonconforming special cause failures that are of most interest. Therefore, on the surface, using Poisson to find special cause failures that do not conform to the requirements of Poisson use may appear to be inconsistent. But here the Poisson is used in reverse to identify data that appear not to conform to Poisson distribution requirements.

The Poisson probability distribution of events is:

$$P(x; \mu) = (e^{-\mu})(\mu^{x})/x! \qquad (1)$$

Where:

e: An approximately 2.71828 constant, the base of the natural logarithm system.

μ: The mean number of events that occur in a specified time period.

x: Specific number of events that occur in a specified time period.

P(x; μ): The Poisson probability that exactly x events are experienced, given the mean is μ.

The general Poisson expression is now adapted specifically to failure events, the mean of which comes from equipment failure dates and the resulting TBF values.

$$\mu = t/MTBF \qquad (2)$$

Where:

MTBF = mean-time-between-failure

t = the specified time period, a TBF value of interest or sum of one or more consecutive TBF values.

So the Poisson distribution for failure events gives the probability of any specific number of failures x and is dependent on the time interval and MTBF, as below:

$$P(x; t/MTBF) = (e^{-t/MTBF})(t/MTBF)^{x}/x! \qquad (3)$$

Figure 1 is the Poisson distribution for a particular data set discussed by Ascher and Feingold [1]. They discuss that an obvious change following an overhaul went unnoticed for 16 years although this data were almost certainly the most widely cited in repairable system literature. This data set will be used throughout the paper.

3 CONTEMPORANEOUS DATA ANALYSIS

In this section, we demonstrate the feasibility of a detailed analysis immediately upon each failure event. This proactive analysis method gets maximum value from failure time data.

Currently it is customary to analyze data if there is an apparent trend or a reliability issue is otherwise recognized. The cost of these infrequent analyses may not be a significant factor. With contemporaneous analysis, all failures on all monitored equipment are analyzed before there is an obvious reason to do so. The analysis results flag significant events for human eyes to be directed to the situation. This high analysis volume requires the process to be as automated as possible, although the methodologies are certainly applicable to ad hoc analysis of single equipment data sets.

Fortunately, today most data are electronic and available in a form that can be made useable. A semi-automated process currently in use is outside the scope of this paper, but is mentioned because the trending and analysis methodologies were, in part, driven by the automation requirement. The trending and analysis methods are discussed below. Again, these methods are appropriate for individual one-at-a-time use.

3.1 Residuals Trending

Figure 1 data are trended in figure 2 in a way to most sensitively see trends. The ordinate is cumulative time (cumulative TBF) and the abscissa is the difference between the failure count predicted by the data set MTBF and the actual failure count. This trend method allows ready selection of the TBF data to be assessed and the generating p-values. In figure 2, the time of last failure and time of the 10^{th} failure back are inputs. All other inputs for equation 3 are then calculated. The lookback calculated p-value is here called p-v1 for a time period equal to the last failure TBF. This p-value is the probability of 1 or more failures within a time period equal to the TBF of the last failure. For a look back of 10 failures where a trend started, the p-value is the probability of 10 or more failures occurring within a time period equal to the sum of the last 10 TBF values. This value is called p-v10.

The Poisson p-v10 value for 10 failures within the time period of 350 days (the sum of the last 10 TBF values) when the MTBF is 95.7 (the MTBF at the 23^{rd} failure) is 0.0045. Therefore, this group of 10 failures are unlikely to have occurred by random chance. A problem has been statistically proven, but unnecessary failures have occurred that could have been avoided. The object of data analysis is, or should be, earliest possible recognition and initiation of action to intervene in the reliability degradation.

To see the earliest possible problem indication, we go back in time ignoring the last 7 failures. This is done by deleting the last 7 TBF values in figure 2. With contemporaneous data at the time of the 16th failure, the p-value for the last three failures (failure 16, 15 and 14) is 0.0325. In keeping with the established terminology, p-v3 at the 16th failure is 0.0325. This could trigger a response, depending upon the value of asset uptime, repair cost, etc. This probability can be considered the likelihood of a false positive. A false positive is one in which we think there is a special cause when the low p-value is due only to randomness is a HPP. The risk of investigating a common cause, random failure can be balanced

TBF	
413	118
14	overhaul
58	34
37	31
100	18
65	18
9	67
169	57
447	62
184	7
36	22
201	34

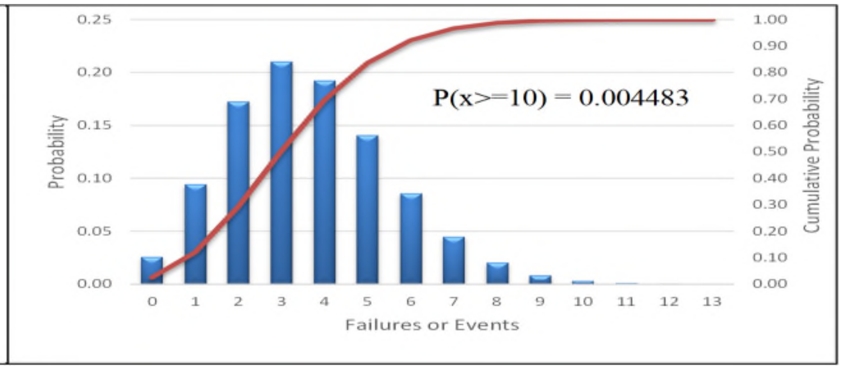
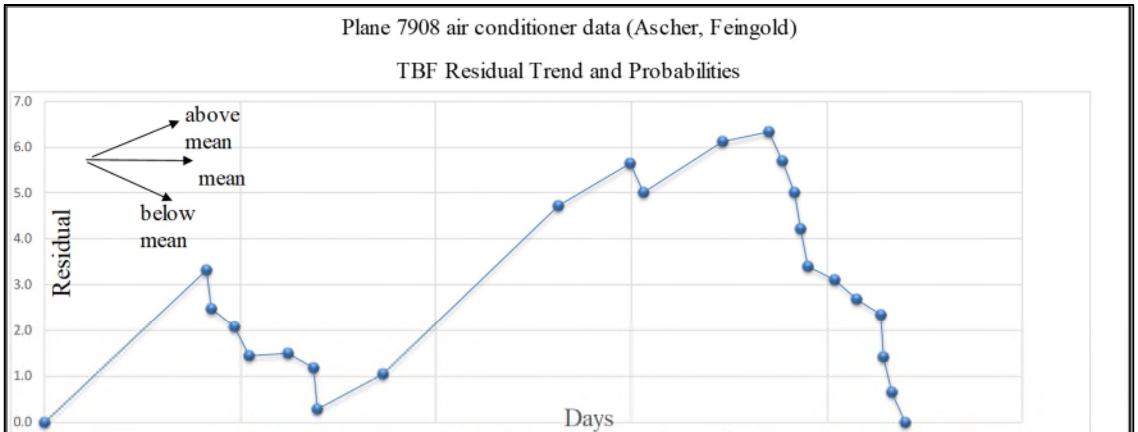


Figure 1 - The probability of 10 or more failures, within a time equal to the sum of the last 10 failures when the MTBF is that of the data set. The low p-value (called p-v10) is strong evidence that the 10 failure times do not belong to the data set null distribution. There is a special cause for failure worthy of investigation. For this data, it is known that an overhaul occurred just prior to the last 10 failures.



			Time unit =	Days				
Between Any Two Failure Events			MTBF for	350	Long term MTBF For	2201	Days	
Enter an x1 value:	1851	Days	=	35	=	95.70	Days	P-values
Enter an x2 value:	2201	Days	Probability that count =	10	within interval of	350	=	0.003045
Time between events:	350	Days	Probability that count <	10	within interval of	350	=	0.995517
Failures between events:	10	count	Probability that count <=	10	within interval of	350	=	0.998562
			Probability that count >	10	within interval of	350	=	0.001438
			Probability that count >=	10	within interval of	350	=	0.004483

Figure 2 - Residuals trend of figure 1 data. The trend is sensitive and allows data of interest to be selected for p-values. The lower right p-value is identical to figure 1 and is found with related Poisson look-up tables in Excel. It is the probability of 10 or more failures in a time interval of 350 days when the MTBF is 95.7

Failure	TBF	MTBF	p-v1	p-v2	p-v3	p-v4	p-v5	p-v6	p-v7	p-v8	p-v9	p-v10	p-v11	p-v12
13	118	142.38	0.5634	0.6552	0.4545	0.5235	0.8200	0.8188	0.7075	0.6312	0.5875	0.4903	0.4171	0.3146
14	34	134.64	0.2232	0.3115	0.4870	0.3279	0.4210	0.7667	0.7774	0.6640	0.5933	0.5574	0.4666	0.4003
15	31	127.73	0.2155	0.0929	0.1745	0.3542	0.2352	0.3365	0.7137	0.7366	0.6230	0.5583	0.5297	0.4450
16	18	120.88	0.1384	0.0630	0.0325	0.0877	0.2421	0.1591	0.2594	0.6575	0.6938	0.5810	0.5226	0.5015
17	18	114.82	0.1451	0.0400	0.0215	0.0125	0.0447	0.1639	0.1077	0.1997	0.6041	0.6528	0.5421	0.4900
18	67	112.17	0.4497	0.1761	0.0659	0.0333	0.0185	0.0454	0.1492	0.1006	0.1856	0.5776	0.6309	0.5241
19	57	109.26	0.4065	0.3137	0.1428	0.0612	0.0328	0.0188	0.0410	0.1312	0.0901	0.1687	0.5500	0.6083
20	62	106.90	0.4401	0.3058	0.2534	0.1267	0.0598	0.0337	0.0200	0.0395	0.1205	0.0843	0.1580	0.5279
21	7	102.14	0.0662	0.1473	0.1279	0.1235	0.0588	0.0270	0.0155	0.0095	0.0224	0.0840	0.0590	0.1235
22	22	98.50	0.2002	0.0357	0.0668	0.0660	0.0706	0.0336	0.0156	0.0092	0.0058	0.0150	0.0639	0.0453
23	34	95.70	0.2990	0.1170	0.0293	0.0437	0.0442	0.0492	0.0240	0.0114	0.0069	0.0045	0.0118	0.0529

Figure 3 - A portion of the spreadsheet probability map for the data of figures 1 and 2 for failures 13 thru 23 with p-values 1 thru 12. The MTBF and p-values are calculated with contemporaneous data, i.e., each line of data uses only history, not future data. The last failure p-v10 (10 failure look back) is identical to figures 1 and 2 values. Conditional formatting shows low p-values starting at failure 16 that could have triggered earlier intervention.

against the need for high reliability. If reliability is critical, we would investigate. In this case study, we know the future. Seven more failures are predestined to occur without recognition and action.

3.2 Poisson Probability Map

A probability map is an effective way to see and measure trends in TBF data. Figure 3 is such a map with only a portion of the map shown because of size. The TBF data are the same as used in figures 1 and 2. The nomenclature for figure 3 is explained by example. P-v1 refers to the probability of 1 or more failures within a time interval equal to the last TBF. P-v10 is the probability of 10 or more failures occurring in the time interval equal to the sum of the last 10 TBF. The first statistical alarm of significance is 0.0325 for p-v3 at failure 16, and how this is obtained will be described as an example. The Excel equation in the cell row with failure number 16 and column p-v3, using values instead of cell locations for clarity, is:

p-v3=[1-POISSON.DIST(2,(18+31+34)/120.88,TRUE)]   (4)

The null hypothesis HPP would produce 3 or more failures within this time (18+31+34) by random chance. Therefore, we conclude with moderate confidence (the complement of 0.0325) that the TBF data indicate a system degradation upon failure number 16. This conclusion is based on contemporaneous data available upon failure 16. There is no knowledge of the next 7 failures. The probability map values are identical to the residuals Excel tool of figure 2, but all of the p-values are seen simultaneously. The increasing strength of evidence of equipment with a new problem, in this case worse than new following overhaul, is evident in figure 3. The p-values in the box show continuously increasing strength of evidence of a problem as failure data arrives. That is clear now but was not seen at the time or for the next many years although looked at often in the reliability literature [1].

4 P-VALUE PROBABILITY DISTRIBUTIONS

The p-values of figure 3 are calculated using the cumulative MTBF at each failure. If this MTBF is true, then the p-values are single-valued and true. Most generally, the TBF should be treated as a sample drawn from a population and MTBF is actually a random variable, thus the p-values are a random variable. The MTBF values in the 3rd column of Figure 3 are only best estimates of the true population MTBF. Using these estimates, the appropriate number of TBF values are generated in Excel using a random number generator. A random number from 0 to 1 is used in the following Excel equation to generate a TBF random variable sample from a HPP.

TBF = -MTBF*LN(RAND())   (5)

To demonstrate, at failure number 16, there were 16 TBF data that produced the mean 120.88 seen in figure 3. So equation 5 uses the mean 120.88 and generates 16 TBF values. These 16 TBF values are averaged to give a MTBF random variable value. Each of these MTBF random variables are used in the Poisson probability map to generate the p-value random variables. With this process repeated numerous times with Monte Carlo simulation in the Excel spreadsheet, probability distributions around all the p-values are obtained. The process above described for the 16th failure is conducted on every failure number in the data base. With spreadsheet add-ins, such as @Risk, these are all done concurrently with a single simulation.

Figure 4 is the 16th failure p-v3 distribution with 90% of the values between 0.0127 and 0.101.

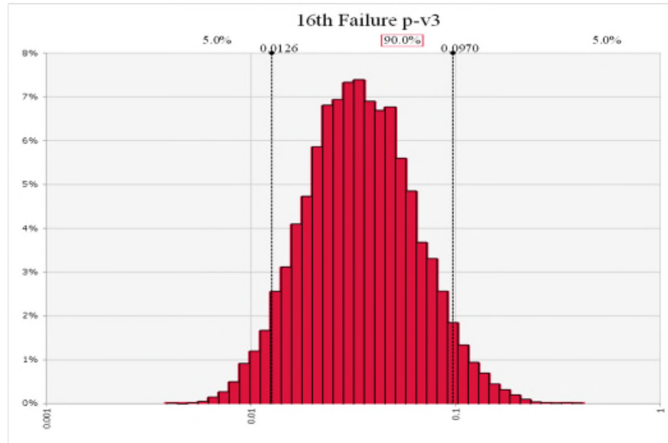


Figure 4 - The 16th failure could have triggered attention. The p-v3 is 0.0325 with only 5% chance of being above 0.1.

5 RESIDUAL MAXIMUM

Residual trends such as figure 2 produce a maximum and minimum value. The most extreme value (largest positive or smallest negative) is the point of maximum change in slope within the data base. A positive value indicates an increasing failure rate (decreasing MTBF) and a negative extreme value is a decreasing failure rate (increasing MTBF). This most extreme residual value is independent of the order of events before and after that data point. It is a data set statistic found to be useful in distinguishing true trends versus HPP random variation. To illustrate, we take the data set at failure number 4. The residual trend is seen in figure 5. The data set extreme value residual at this 4th failure is 2.16 as seen in the figure. Included in figure 5 are five other data sets formed by using the TBF random variable samples generated by equation 5. These additional five data sets are samples from a trendless HPP. The trend indication by sight, but more importantly by the residual statistic, allows the data set trend to be compared to null hypothesis generated sample data sets. With the TBF values used in figure 5 linked to equation 5 TBF random variable samples, one can hit the F9 key and qualitatively judge if random samples tend to exceed the data base maximum (or minimum for reliability growth) even without Monte Carlo. For simplicity, we refer to the method as residual maximum, recognizing it is the smallest negative for reliability growth trends.

The maximum and minimum values for each iteration are captured and the distribution for the 4th failure residual maximum is shown in figure 6. Figure 6 probability distribution is the null hypothesis against which the data set residual maximum can be tested. The p-value is 0.011, indicating only about 1% of HPP trendless random samples have trend indication greater than our data base through failure number 4. So we reject the null hypothesis that there is no trend. With contemporaneous analysis, this would trigger an investigation at failure number 4. Figure 7 is a cumulative distribution for the residual at failure number 2, demonstrating that only two failures are sufficient to test against the null hypothesis.

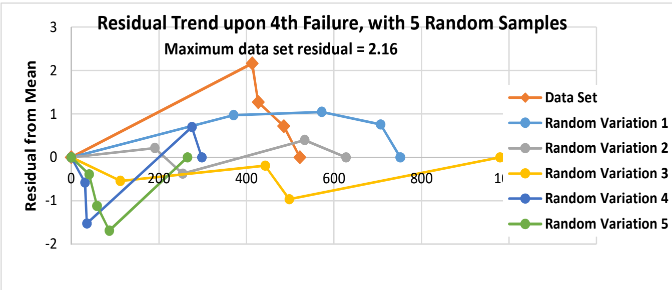


Figure 5 – Residual trend at failure number 4 with 5 data sets using TBF random variable values generated using equation 5. These generated data sets are trendless by definition with apparent trends being random variation.

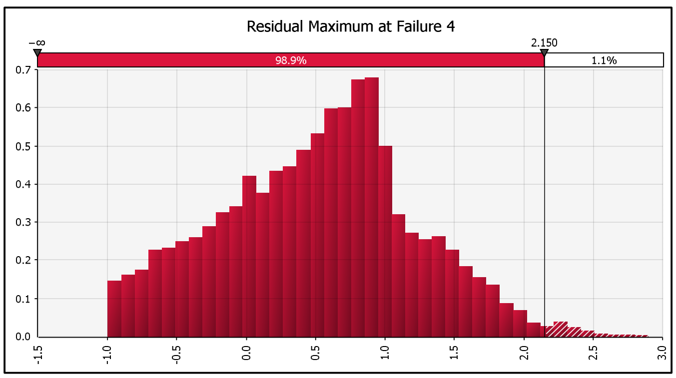


Figure 6 – Probability distribution for the residual maximum at failure number 4. The distribution is the HPP null hypothesis against which the data set residual maximum of 2.16 can be tested. The null is rejected with a p-value of 0.011

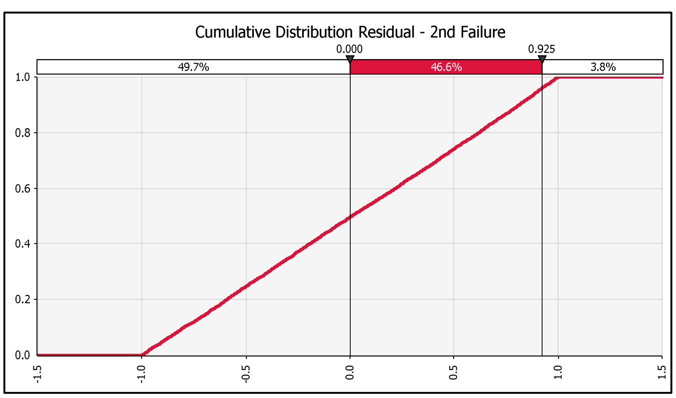


Figure 7 – Cumulative probability distribution for failures 1 and 2 residual. There is only one residual point for each iteration due to only 2 failures. About 3.8% of the random sample residuals are larger than the data set residual of 0.93. The residual maximum method for trend detection is applicable to data sets as small as two.

6 COMPARISON OF METHODS

6.1 Background

Wang and Coit [2] evaluated several trend tests and reported the Crow/AMSAA to be the most robust. Therefore, the Crow/AMSAA was selected for comparison with the Poisson and residual maximum methods. The data base was analyzed by the three methods emulating contemporaneous analysis, that is, at failure number 2 for example, only data for the first two failure are used. The analysis is blind to any future failures. (This is the protocol used throughout).

6.2 Results

Table 2 presents the evaluation results. The Crow/AMSAA test for a trend is known to be invalid for small data sets, but the results are included. The exact number at which Crow/AMSAA should not be used may not be universally agreed upon. This area of unsuitability is marked in table 2. The Crow/AMSAA test was conducted as described by Wang and Coit [2] and is identical to that described by Crow much earlier.

First we considering the last 10 failures identified as being a trend in figures 1, 2 and 3. The Crow/AMSAA identified a trend at failure number 22 with a p-value of 0.0127. The Poisson was more sensitive by identifying a trend earlier at failure number 17 with a p-value of 0.0125 for the p-v4 (a lookback of 4 failures), about the same p-value as the Crow/AMSAA at failure number 22. The 90% confidence values from the p-v4 probability distribution are 0.0036 and 0.0515. The residual maximum did not detect the trend until failure number 23 with a p-value of 0.014, a comparable p-value. So the performance for the trend of the last 10 failure is as follows:

- Poisson detected at failure 17
- Crow/AMSAA detected at failure 22
- Residual maximum detected at failure 23

The first trend in the data set is actually immediately following the first failure. This trend was not mentioned by Ascher and Feingold [1]. Crow/AMSAA is not valid for this small data set. The Poisson had moderately high p-values in this region as well as broad probability distributions. However, the residual maximum p-value at failure number 4 is 0.0110. The residual maximum method is uniquely able to detect trends in very small data sets –as small as two failures. So for the first trend the performance summary is as follows:

- Residual maximum detected at failure 4
- Poisson p-values too high to detect
- Crow/AMSAA invalid in this region

The comparison results indicate the Poisson and Residual maximum methodologies are competitive with other methods and better in some instances. Their value is in testing against the null hypothesis of no trend. Their efficiency makes large scale data analysis practical.

Ascher, Feingold Data Set		Crow/AMSAA			Poisson				Residuals	
		Beta and H_o rejection p-value			p-values		90% confidence from Monte Carlo distributions		Maximum/Minimum Residual p-values from Monte Carlo distributions	
Failure number	TBF	Beta	2N/Beta	Chi square p-value	Specific minimum p-value	minimum p-value magnitude	5% p-value	95% p-value	Max/Min Residual p-value	Trend direction
1	413									
2	14	59.9944	0.0667	0.0005	p-v1	0.0635	0.0271	0.2970	0.0390	degradation
3	58	10.4143	0.5761	0.0032	p-v2	0.0741	0.0197	0.4840	0.0220	degradation
4	37	7.8644	1.0172	0.0019	p-v3	0.0526	0.0100	0.4520	0.0110	degradation
5	100	4.1332	2.4194	0.0080	p-v4	0.0902	0.0136	0.6140	0.0180	degradation
6	65	3.5156	3.4134	0.0081	p-v5	0.0950	0.0127	0.6330	0.0200	degradation
7	9	3.9221	3.5696	0.0025	p-v1	0.0692	0.0079	0.5850	0.0100	degradation
8	169	2.4195	6.6129	0.0200	p-v7	0.1303	0.0159	0.7480	0.0450	degradation
9	447	1.3556	13.2781	0.2252	p-v8	0.2793	0.0432	0.9020	0.1510	growth
10	184	1.2787	15.6405	0.2613	p-v9	0.3026	0.0458	0.9130	0.1540	growth
11	36	1.3651	16.1161	0.1899	p-v1	0.2278	0.1550	0.3720	0.2660	growth
12	201	1.2747	18.8282	0.2389	p-v11	0.3108	0.0440	0.9170	0.2490	growth
13	118	1.2739	20.4092	0.2282	p-v12	0.3146	0.0410	0.9190	0.3060	degradation
14	34	1.3408	20.8824	0.1698	p-v1	0.2232	0.1570	0.3410	0.3450	degradation
15	31	1.4059	21.3391	0.1230	p-v2	0.0929	0.0480	0.2020	0.3220	degradation
16	18	1.4801	21.6197	0.0828	p-v3	0.0325	0.0125	0.0988	0.3060	degradation
17	18	1.5514	21.9161	0.0544	p-v4	0.0125	0.0036	0.0515	0.1820	degradation
18	67	1.5609	23.0635	0.0467	p-v5	0.0185	0.0046	0.0869	0.1450	degradation
19	57	1.5790	24.0658	0.0383	p-v6	0.0188	0.0041	0.0982	0.1060	degradation
20	62	1.5883	25.1841	0.0326	p-v7	0.0200	0.0037	0.1130	0.0830	degradation
21	7	1.6591	25.3148	0.0195	p-v8	0.0095	0.0014	0.0722	0.0430	degradation
22	22	1.7092	25.7434	0.0127	p-v9	0.0058	0.0007	0.0511	0.0220	degradation
23	34	1.7406	26.4284	0.0091	p-v10	0.0045	0.0005	0.0446	0.0140	degradation

Table 2 - Crow/AMSAA, Poisson and Residual methods compared using the Ascher/Feingold data set. Crow/AMSAA is known to be invalid for very small data sets but values are shown for completion, but marked. Smaller p-values are shaded for ready comparison.

REFERENCES

1. H. Ascher, H. Feingold, *Repairable Systems Reliability – modeling, inference, misconceptions and their causes*, New York, Marcel Dekker, 1984.
2. Peng Wang, D. W. Coit, "Repairable Systems Reliability Trend Tests and Evaluations", *Proc. Ann. Reliability & Maintainability Symp*., (Jan.) 2005.

BIOGRAPHIES

Jan B. Smith, Senior Reliability Consultant for Zachry Industrial, Inc. has experience in root cause failure analysis; finite element analysis; statistical and probabilistic analysis; vibration analysis; RCM; and capacity and availability forecasting. He has chaired conferences on plant reliability, authored technical papers on the subject, and has organized and taught in house and public seminars on root cause analysis. He is a registered Professional Engineer. smithjb@zachrygroup.com

Oliver A. Sac received a BSME from the University of Houston and is currently a Reliability Engineer for Chevron Phillips Cedar Bayou. He is responsible for rotating equipment reliability in a large multiple processing unit petrochemical plant including ethylene, utilities, and olefins units. His responsibilities include equipment root cause failure analysis and predictive maintenance, developing planned maintenance plans, and machinery improvements and redesign. His work experience includes applying probability methods to equipment reliability data. sacoa@cpchem.com

Kevin Bordelon, P.E., SMRP, has BSChE and MBA degrees with experience in the chemical, petrochemical and refining industries. His background includes process engineering, production and plant management, and reliability management. He is Director of Maintenance and Reliability Services for Zachry Industrial, Inc. with responsible for reliability and maintenance technical support. He also serves as Refinery Business Unit General Manager supporting the refining industry.
bordelonk@zachrygroup.com