ISSN 0511-5728
The West Indian Journal of Engineering
Vol.35, No.1, July 2012, pp.4-9

A Review on Trend Tests for Failure Data Analysis

Albert H.C. Tsang

Department of Industrial and Systems Engineering, The Hong Kong Polytechnic University, Hung Hom, Hong Kong, China;
E-mail: albert.tsang@polyu.edu.hk

(Received 21 June 2011; Revised 22 January 2012; Accepted 06 March 2012)

Abstract: *Trend detection is an important task in failure data analysis in both reliability and maintenance. Usually, detecting certain possible trend in failure events of systems or items is the first step of data handling, and the result of trend detection indicates some directions for further statistical analysis of these data. Available statistical tests commonly used for detecting trends in failure events collected over time as well as the underlying hypotheses being tested are reviewed in this paper. Directions for future research on statistical trend tests are suggested at the end of the paper. These suggestions will address issues concerning analysis of failure data obtained from single and multiple systems.*

Keywords: *Failure data analysis, statistical trend tests, homogeneous Poisson process, monotonic trend, multiple systems*

1. Introduction

Reliability evaluation and maintenance decision modeling typically involve characterization of a physical asset's lifetime distribution. A range of techniques, such as Weibull analysis, are available to estimate an item's lifetime distribution. These procedures are developed on the premise that the process generating the failure events is stable. That means, statistically speaking, all the failure times observed are independently and identically distributed (iid). In reality, this condition may not apply. For example, intervals between failures observed in a repairable system may show a tendency to decrease with time due to the cumulative effect of imperfect repairs in successive repair cycles. A repair is imperfect if it simply returns the item to an operational state instead of the "as-good-as-new" condition.

In another scenario, these intervals may have shown an increasing trend that could be the effect of incremental design improvement introduced in successive repair cycles. It is a common practice to model a repairable item's time to failure as a stationary distribution of its usage since the last repair action, i.e., the distribution is invariant from one repair cycle to another. However, this approach to modeling will be inappropriate if the failure time data are generated from an unstable process. Thus, it is necessary to detect trends in the rate of occurrence of failures before any attempt is made to characterize an item's lifetime distribution as a stationary process.

This paper reviews the commonly used graphical techniques and statistical tests for detecting trends in sequence of failure events, and the underlying hypotheses being tested by these statistical procedures. Practical issues of applying statistical trend tests on single as well as multiple systems are also discussed. Directions for future research on statistical trend tests that will address these practical issues are suggested at the end of the paper.

2. Trend Tests

2.1 Graphical techniques

Simple graphical techniques are used to help determine whether the reliability of a system is improving or deteriorating. These techniques are particularly useful for identifying the salient features of the data and for checking the assumptions made in fitting formal models to data. (Ascher and Feingold, 1984)

(a) Plotting cumulative failures versus cumulative time on linear paper

A trend of increasing inter-arrival time of failure events in a system indicates an improving system; a trend of decreasing inter-arrival time of failure events in a system indicates a deteriorating system.

(b) Estimating average ROCOF in successive time periods

Rate of occurrence of failure (ROCOF) is the rate of arrivals of failure events occurring at a particular time of operation defined as follows:

$$\hat{v}(t) = \frac{N(t_i) - N(t_{i-1})}{t_i - t_{i-1}}$$

In the above formula, $t_{i-1} < t < t_i$ and $N(t_i)$ is the total number of failures observed from $T = 0$ to the end of the i^{th} interval.

(c) Duane plots

Cumulative mean-time-between-failures (CMTBF) is defined as follows:

$$\widehat{\text{CMTBF}} \equiv \frac{t}{N(t)}$$

where $N(t)$ is the total number of failures observed from $T = 0$ to the end of the observation period ($T = t$).

Duane observed that a plot of the CMTBF versus cumulative operating (or test) time on log-log paper is linear, and the resulting graph is known as a Duane plot (O'Connor, 2002). This plotting technique is typically applied to analyze failure data obtained from product development tests conducted to expose design weaknesses, and design improvements are introduced to address these weaknesses before the product is resubmitted for further testing. Thus, reliability often grows with the testing program. A Duane plot with a positive slope demonstrates reliability growth. A larger slope indicates faster rate of change of reliability after each design improvement.

(d) TTT Plot (Kvaløy and Lindqvist, 1998)

The Total Time on Test (TTT) plot is the most well known graphical technique. Barlow and Davis (1977) discussed a TTT plot for data from repairable systems based on the NHPP (non-homogeneous Poisson Process) model. The plot is based on the n independent NHPP's with common intensity function $\lambda(t)$ is observed, and all observations are contained in the time interval [0, T].

The NHPP differs from the HPP only in that the rate of occurrence varies with time rather than being a constant. t_i represents the time of the i^{th} arrival in the superposed process; $p(u)$ represents the number of processes under observation at time u.

$\mathrm{T}(t) = \int_0^t p(u)du$ denotes the total time on test from time 0 to time t.

The scaled TTT plot for NHPP's is a plot of the scaled total time on test statistic, i.e.,

$$\frac{T_i}{T_n} = \frac{\int_0^{T_i} p(u)du}{\int_0^{T_n} p(u)du}$$

versus scaled failure number $\frac{i}{n}$ (on the abscissa).

"No trend" corresponds to a TTT plot located near the main diagonal of the plot. The shape of the TTT plot indicates the type of trend that exists in the failure data (see Figure 1).

The graphical techniques reviewed in this section are mainly applied in practice to detect possible trends in a time series of failure events. While they enable visual judgment of trends, these graphical techniques do not provide statistical evidence to cast doubt on the "no-trend exists hypothesis", i.e., the null hypothesis.

2.2 Statistical trend tests

In failure data analysis, another category of trend tests covers the statistical tests for null hypotheses against their alternatives. Table 1 shows the taxonomy of these trend tests as proposed by Ascher and Feingold (1984).

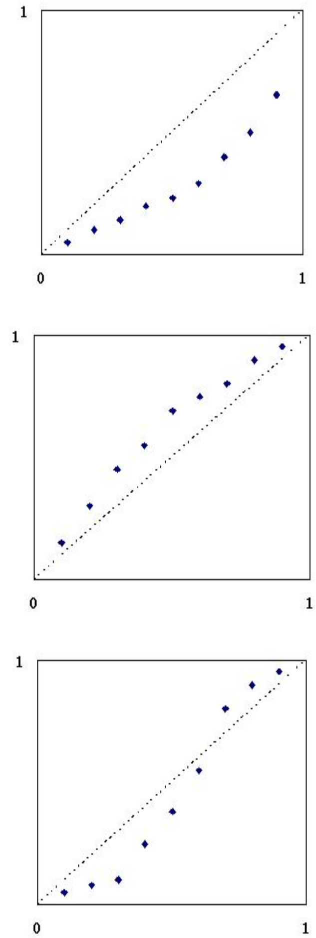

Figure 1. TTT plots from NHPP's with decreasing, increasing and bathtub shaped intensity function (from top), respectively.

Table 1 Taxonomy of trend tests

Null Hypothesis	Alternate Hypothesis
HPP (homogeneous Poisson process) (H)	Monotonic trend (M)
Renewal process (R)	Non-monotonic trend (N)
General: Stationary sequence (S)	

Let X_i be the inter-arrival time between the $(i-1)$-th and the i-th failures. A sequence of failure events is a homogeneous Poisson process (HPP) if the X_i's are independent and identically exponentially distributed, i.e., the ROCOF is constant. A non-homogeneous Poisson process (NHPP) differs from HPP in that the ROCOF varies with time.

Let $F_{X_i}(x)$ be the cumulative distribution function of X_i. A sequence of failure events exhibits monotonic trend if it satisfies the condition $F_{X_i}(x) > (<) F_{X_j}(x)$, for every $\boldsymbol{i} \geq 1$, every $\boldsymbol{j} > \boldsymbol{i}$, and every $\boldsymbol{x} > 0$, X_i and X_j are chronologically ordered but independent random variables. A sequence of failure events is said to exhibit non-monotonic trend if there is a tendency for successive X_i 's to decrease or increase even though the inequality $F_{X_i}(x_i) > (<) F_{X_j}(x_j)$ is not fully met.

A renewal process is a generalization of HPP. That is, the X_i's are independent and identically distributed (IID). Since X_i is not restricted to be exponentially distributed, the hazard rate, $h_{X_i}(x_i)$, in general is a function of x_i. When $h_{X_i}(0) < (>) h_{X_i}(x_i)$ for all $x_i > 0$, the system is said to be "good (bad)-as-new" after each repair.

A sequence of X_i's is said to be stationary if the joint distribution of any j of the inter-arrival times is invariant under a jump from one failure event to another, for any j ≥ 2.

Most of the trend tests are designed for categories H-M and R-M in the Ascher and Feingold Taxonomy. These tests are reviewed as follows.

(a) Category H-M

Tests proposed for discriminating between a HPP, or H in short, and a process with monotonic trend include Laplace (Bartholomew, 1955) and MIL-HDBK-189. These two tests are introduced below.

Laplace's test

Figure 2 shows the time series of failure events (indicated as dots on the time line) observed in a time-terminated test.

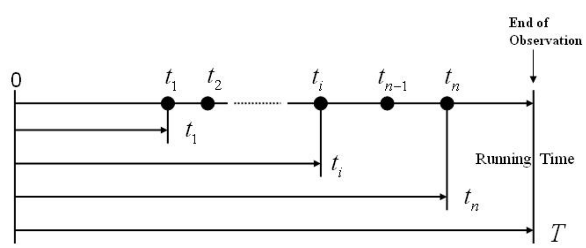

Figure 2. Time-terminated Laplace Test

The test statistic for the Laplace trend test applied to time-terminated data is:

$$U = \sqrt{12N(t_n)}\left[\frac{\sum_1^n t_i}{T \bullet N(t_n)} - 0.5\right]$$

t_i denotes the running time of a repairable item at its i^{th} failure, where i = 1, …, n. Let $N(t_n)$ be the total number of failures observed up to time t_n, and the observation terminates at time T when the item is not in the fail state.

When the observation terminates at a failure event, the test statistic for the trend test applied to such failure-terminated data is:

$$U = \sqrt{12N(t_{n-1})}\left[\frac{\sum_1^{n-1} t_i}{t_n \bullet N(t_{n-1})} - 0.5\right]$$

U is normally distributed with mean = 0 and standard deviation = 1 if the inter-arrival times of failure events are generated from a HPP. When U is significantly small (negative), the null hypothesis of HHP is rejected, indicating evidence of reliability growth; when U is significantly large (positive), the null hypothesis of HPP is rejected as well, indicating evidence of reliability deterioration (Jardine and Tsang, 2006).

Suppose the significance level, α, of the test is set at 5%, the lower and upper bounds of the test statistic for a two-sided test are −1.96 and 1.96, respectively. If the U value is within this range, a Poisson model can be used to characterize the inter-arrival times, X_i 's, of the observed failure events.

Laplace's test is optimal against the alternate hypothesis of non-homogeneous Poisson process (NHPP) in which:

$$\rho_1(t) = \exp(\alpha_0 + \alpha_1 t)$$

$\rho_1(t)$ is the ROCOF, also known as the peril rate, of the NHPP. α_0, α_1 are non-negative parameters.

MIL-HDBK-189 test (1981)

This test is for tracking reliability growth developed by the US Army Materiel Systems Analysis Activity (AMSAA) (Unkle and Venkataraman, 2002). It assumes that the ROCOF of failure events has a peril rate of $\rho_2(t) = \lambda\beta t^{\beta-1}$ (Crow, 1974). λ is a non-negative parameter, and β is known as the growth parameter.

The hypotheses to be tested are:

$H_0 : \beta = 1$, i.e. the failure events are generated from a HPP, or $\rho_2(t) = \lambda$

$H_1 : \beta \neq 1$ (NHPP)

In this test, the test statistic for failure data obtained from a time truncated test is:

$\chi^2_{2n} = \dfrac{2n}{\hat{\beta}}$, where the growth parameter is

$$\hat{\beta} = \frac{n}{\sum_{i=1}^{n} \ln\left(\frac{T}{t_i}\right)}.$$

$t_1 < t_2 < \ldots < t_n < T$. n denotes the total number of failure observations.
Under the null hypothesis of HPP (no growth), χ^2_{2n} has a Chi-square distribution with $2n$ degrees of freedom.
The test statistic for the failure truncated test is

$\chi^2_{2(n-1)} = \frac{2n}{\hat{\beta}}$, where the growth parameter is

$$\hat{\beta} = \frac{n}{\sum_{i=1}^{n-1} \ln\left(\frac{t_n}{t_i}\right)}$$

This test statistic is Chi-square distributed with $2(n-1)$ degrees of freedom when the null hypothesis of HPP is true.

The null hypothesis is rejected for small (large) values indicating an improving (deteriorating) system. The statistic $\hat{\beta}$ estimates the growth parameter β. In the case of no growth, β is equal to 1; when $\beta < 1$, the process indicates reliability growth, when $\beta > 1$, the process indicates reliability deterioration.

(b) Category R-M

The null hypothesis in this category is more general than the one of HPP, and its tests are designed to distinguish between a renewal process and a monotonic trend.

The Mann Reverse Arrangements Test (Mann, 1945)

This is a non-parametric test that does not require any assumption about the failure process, other than being a renewal process. By counting the reverse arrangements among the chronologically ordered independent inter-arrival times X_1, X_2..., X_n. a reverse arrangement is said to exist whenever $X_i < X_j$ for $i < j$.

Define $\Re_m$ as the total number of reverse arrangements for $I = 1, \ldots, n-1$ and $j = 2, \ldots, n$. $\Re_m$ is calculated by comparing every inter-arrival time with every later inter-arrival time. Since a total of $n(n-1)/2$ comparisons are to be made,

$$\mathrm{E}\left(\Re_n \middle| \text{renewal}\right) = \frac{(n-1)n}{4}$$

Under the null hypothesis of renewal, there will be no tendency for earlier inter-arrival times to be less than, or greater than, later ones.
The variance of $\Re_m$ was

$$\mathrm{Var}\left(\Re_n \middle| \text{renewal}\right) = \frac{(2n+5)(n-1)n}{72}$$

Mann showed that under the renewal hypothesis, $\Re_n$ is approximately normally distributed for $n \geq 10$ and tabulated $\Pr\left(\Re_n \leq \Re_n^*\right)$ for $\Re_n = 0, 1, \ldots, n(n-1)/2$. The one sided probability of obtaining $\Re_n$ or less reverse arrangements with n inter-arrival times can be obtained from Mann's Table (Mann, 1945).

The Pair-wise Comparison Non-parametric Test (PCNT)

This is a modified version of the Mann Reverse Arrangement Test proposed by Wang and Coit (2005). The test statistic is:

$$U_p = \frac{\Re_n - n(n-1)/4}{\sqrt{\frac{(2n+5)(n-1)n}{72}}}$$

$\Re_n$ is as defined in the Mann Reverse Arrangement Test. The null hypothesis that the observed events are generated from a renewal process is not rejected if $-z_{\alpha/2} \leq U_p \leq z_{\alpha/2}$ where α is the significance level of the test.

The Lewis-Robinson Test (Lewis and Robinson, 1974)

This test is a modification of the Laplace's test, with the test statistic

$$U_{LR} = \frac{U}{\hat{CV}(X)}$$

U is the Laplace test statistic, and $\hat{CV}(X)$ is an estimate of the coefficient of variation of the X_i's (the inter-arrival time of the i-th failure event).

3. Trend Test for Data from Multiple Systems

Cox and Lewis (1966) proposed a trend test for failure data collected from multiple systems. Suppose multiple independent series of failure observations are available, in which $t', t'', \ldots\ldots,$ denote the arrival times of the i-th failure in the various series, $n', n'', \ldots$ denote the numbers of events observed in the different series, the observations on the various series are terminated at $t', t'', \ldots\ldots,$ respectively, none of which is the time of a failure event.

The hypothesis of this test is that the rates of occurrence, $\rho(t)$, of events observed in all the time series have trends that follow the same model: $\rho(t) = \exp(\alpha + \beta t)$.

The test statistic for the null hypothesis of no trend, i.e., $\beta = 0$ is:

$$U = \frac{\left(\sum t'_i + \sum t''_i + \cdots\right) - \frac{1}{2}\left(n't' + n''t'' + \cdots\right)}{\sqrt{\frac{\left(n't'^2 + n''t''^2 + \cdots\right)}{12}}}$$

When the null hypothesis of no trend is true, U is approximately normally distributed with zero mean and unit variance. When U is significantly larger than 0, there is evidence that the rate of occurrence of failures in some or all the series may increase with time, i.e., $\beta > 0$ in some of the models for $\rho(t)$.

The value of β in the model for $\rho(t)$ that applies to failure observations on a single series can be estimated by solving the following equation (Cox and Lewis, 1966):

$$\frac{n}{\beta} - \frac{nt}{1-\exp(-\beta t)} + \sum t_i = 0$$

This estimate of β is approximately normally distributed with a standard error of:

$$\sqrt{n\left(\frac{1}{\beta} - \frac{t^2 e - \beta\, t}{\left(1-\exp(-\beta\, t)\right)^2}\right)}$$

Applying the above results to test the null hypothesis of $\beta = \beta_0$, in the case of failure observations on a single series, the test statistic is:

$$U = \frac{\frac{n}{\beta_0} - \frac{nt}{1-\exp(-\beta_0 t)} + \sum t_i}{\sqrt{n\left(\frac{1}{\beta_0} - \frac{t^2 e - \beta_0 t}{\left(1-\exp(-\beta_0 t)\right)^2}\right)}}$$

The null hypothesis that the observed events are generated from a process with $\beta = \beta_0$ in its peril rate is not rejected if $-z_{\alpha/2} \le U \le z_{\alpha/2}$ where α is the significance level of the test.

4. Discussion

4.1 Graphical Versus Statistical Trend Tests

Graphical methods tend to mask local variations, even when the sample size is relatively large. It is difficult to recognize the trends due to random variations in the plot. Furthermore, these techniques cannot be used to estimate the confidence interval of the trend estimate; in other words, the trend identified from these techniques is not quantified statistically.

Statistical trend tests, as compared to graphical techniques, place more emphasis on testing significance. The literature review presented in this paper indicates that most of prior studies in the theory of trend testing were published decades ago. Recent publications on trend testing are on applications of trend tests reviewed in this paper, such as Musa (1996). The power for detecting very specific types of trend or for detecting a range of alternative hypotheses should be a criterion for selection of the statistical test to be used. More research on the power of trend tests will help reliability practitioners in making informed decisions on selection of statistical trend tests.

4.2 Statistical trend tests for data from multiple systems

It is noted that there are very few papers on statistical tests for detection of trends in failure data observed on multiple systems, an issue commonly encountered in industry. Two categories of multiple systems commonly encountered in industry are:

(a) ***Failure data from similar items with a Serial Number***

Items are manufactured and delivered to customers. Suppose the first failure of each item is covered by warranty. Thus, the times to first failure of these items will be known when they are returned to the customer service center for repair. A fully functional computerized maintenance management system (CMMS) that allows for maintenance information on each item to be recorded will facilitate the management of such data.

The sequence of manufacture of the returned items can be tracked if their serial numbers are known. However, missing data often exist in serial number problems because some customers may not ask for service when the purchased item fails. Future research is therefore proposed to develop trend tests that can be applied to handle serial number problems with missing data.

(b) ***Failure data from identical multiple systems***

Systems are considered identical when their design specifications as well as their working environments are similar. However, situations commonly exist in practice may cast doubts on whether these systems are indeed identical, such as:

1) Design reviews on the system may result in design improvements. Hence, systems manufactured in different periods may not be identical in reliability performance.
2) Multiple systems of the same design are used in different working environments, such as drive motors used in mining equipment may also be used in escalators of office buildings. It is difficult to judge that these systems are identical without testing.
3) Imperfect repair(s) made on systems in the group.

Proschan (1963) suggests a means of pooling test results by combining the test results for two or more systems. Except for the case with HPP as null hypothesis, it is not appropriate to pool failure times from multiple systems because the probabilistic model for the pooled data is unknown, except in asymptotic cases. Even under the HPP hypothesis, it is advisable to pool test statistics, rather than failure times, because of the lack of robustness against other null hypotheses (e.g., Lewis and Cox (1966) and Clifford (1982)).

Lewis and Cox (1966) argue that if a pooled test is required, it is best to take as null hypothesis that the series individually are simple Poisson processes, possibly with different peril rates for different series. They used the technique introduced in Section 3 of this paper to analyze the failure data of air-conditioners in a number of aircraft. After applying the Laplace's trend test on failure data of air-conditioner in individual aircraft, failure data from air-conditioners in multiple aircraft were pooled to test the null hypothesis that failure data for each air-conditioner follows an exponential distribution, i.e., $\beta = 0$. Procedures for estimating the β values for individual series of failure events and for pooled data, as well as those for determination of confidence interval for the β estimates were also demonstrated.

Since failure data may not be exponentially distributed in practice, more general methods for trend detection in data sets obtained from multiple systems are desired. Extension of the method proposed by Cox and Lewis (1966) could be a solution for such applications.

5. Conclusion

Graphical techniques and statistical tests for detection of trends in failure data analysis are reviewed. The currently available trend tests in common use are adequate for testing hypotheses in categories H-M and R-M of the Ascher-Feingold Taxonomy.

Research agenda for trend tests has been identified with a view to providing reliability practitioners with useful procedures for selection of statistical tests that are sensitive to detect specific types of trend. Future research efforts that focus on trend tests of failure data obtained from multiple systems are also recommended.

Acknowledgment

The author would like to express his sincere thanks to the Research Committee of The Hong Kong Polytechnic University for financial support of the research work presented in this paper (Project No. G-U379).

References:

Ascher, H. and Feingold, H. (1984), *Repairable Systems Reliability, Modeling, Inference, Misconceptions and Their Causes*, Marcel Dekker, Inc., p.74, 78

Barlow, R.E. and Davis, B. (1977), "Analysis of Time Between Failure for Repairable Components", In: Fussell, J.B. and Burdick, G.R. (eds), *Nuclear Systems reliability Engineering and Risk Assessment*. SIAM, Philadelphia, pp.543-561

Bartholomew, D.J. (1955), "Discussion of Cox (1955)", *Journal of the Royal Statistical Society. Series B (Methodological)*, Vol.17 No.2, pp.162-163.

Bates, G.E. (1955), "Joint distributions of time intervals for the occurrence of successive accidents in a generalized Polya scheme", *The Annals of Mathematical Statistics*, Vol.26, pp.705-720.

Clifford, P. (1982), "Some general comments on nonidentifiability", In: LeCam, L. and Neyman, J. (eds), *Probability Models and Cancer*, North-Holland, New York, pp.81-81

Cox, D.R. (1955), "Some statistical methods connected with series of events", *Journal of the Royal Statistical Society - Series B (Methodological)*, Vol.17, No.2, pp.129-164.

Cox, D.R. and Lewis, P.A. (1966), *The Statistical Analysis of Series of Events*, Methuen, London, pp.45-51

Crow, L.H. (1974), "Reliability analysis for complex, repairable systems", *Reliability and Biometry*, SIAM, pp.379-410.

Jardine, A.K.S. and Tsang, A.H.C. (2006), *Maintenance, Replacement, and Reliability: Theory and Applications*, CRC Press, Boca Raton

Kvaløy, J.T. and Lindqvist, B.H. (1998), "TTT-based tests for trend in repairable systems data", *Reliability Engineering and System Safety* Vol.60,pp.13-28

Lewis, P.A. and Robinson D.W. (1974), "Testing for a monotone trend in a modulated renewal process", In: Proschan, F. and Serfling, R.J. (eds), Reliability and Biometry, SIAM, Philadelphia, pp.163-182.

Mann, H.B. (1945), "Nonparametric tests against trend", Econometrika, Vol.13, pp.245-259

MIL-HDBK-189 (1981), *Reliability Growth Management*, Headquarters, U.S. Army Communications Research and Development Command

Musa, J.D. (1996), "Software reliability-engineered testing", *Computer*, Vol.29, No.11, pp.61-68.

O'Connor, P.D. (2002), *Practical Reliability Engineering*, 4th ed. Heyden, London, pp.360-366.

Proschan, F. (1963), "Theoretical explanation of observed decreasing failure rate", *Technometrics*, Vol.5, pp.375-383.

Unkle, R. and Venkataraman, R. (2002), "Relationship between Weibull and AMSAA models in reliability analysis: A case study", *International Journal of Quality and Reliability Management*, Vol.19, Nos.8/9, pp.987-988.

Wang, P. and Coit, D.W. (2005), "Repairable systems reliability trend tests and evaluation", *Proceedings of the Annual Reliability and Maintainability Symposium*, Alexandria, Virginia, USA, Jan. 24-27, pp.416-421

Author's Biographical Notes:

Albert H.C. Tsang *is Senior Teaching Fellow of the Department of Industrial and Systems Engineering at The Hong Kong Polytechnic University. He has provided consultancy and advisory services to enterprises and industry support organizations in manufacturing, logistics, public utilities, healthcare and government sectors on matters related to quality, reliability, maintenance, engineering asset management, performance management and assessment of performance excellence – these are also areas of his research interest. He is the author of 'WeibullSoft', a computer-aided self-learning package on Weibull analysis. Apart from publishing papers in various international refereed journals, he is also the author / co-author of three books on various aspects of engineering asset management, and two books on industrial applications of RFID.*

■