International Journal of Performability Engineering, Vol.8, No.3, May 2012, pp. 321-329.

Prediction of Vehicle Reliability using ANN

B. HARI PRASAD[1*], P. BHATTACHARJEE[2] and A. VENUGOPAL[3]

[1]*Military College of EME, Secunderabad-500 015, India*
[2]*Defence Research and Development Laboratory, Hyderabad -500 058, India*
[3]*National Institute Technology, Warangal -506 004, India*

(Received on March 23, 2010, revised on August 13, 2011 and February 18, 2012)

Abstract: ANNs are usually very effective as computational tools and have found extensive utilisation in solving many complex real-world problems. The attractiveness of ANNs comes from their remarkable information processing characteristics pertinent mainly to nonlinearity, high parallelism, fault and noise tolerance besides its learning and generalisation capabilities. The aim of this paper is to familiarise with ANN-based computing (neuro-computing). The predicted and observed vehicle reliability using trained ANN is very close as compared to Weibull probability distribution. The methodology adopted is demonstrated with the help of a case study which includes collection, sorting and grouping of vehicle failure data. Then distribution parameters are estimated and best fitting probability distribution is identified for predicting vehicle reliability. Subsequently the trained ANN (using SLP model) is used to predict the vehicle reliability. Suitability of a RDBMS (Oracle) for training ANN and predicting vehicle reliability is also presented. The developed methodology has been able to predict reliability of vehicle very close to its observed values.

Keywords: *ANN; reliability; probability distribution; probability plots; Chi-square test; SLP.*

1. Introduction

Success of an organisational mission depends upon a variety of factors. One of the most important factors is the preparedness of the machine or the weapon system to optimal performance level. History is full of the examples where failure of these machines has led to premature aborting of the missions besides loss of precious human lives. It is for this reason that the system hardware ought to have a very high degree of 'inherent' and 'achieved' availability. It is a well established fact that a hundred percent reliable machine is impossible to design and manufacture. Moreover, majority of the system hardware is of mechanical type, which deteriorates at varying rates with its usage, under different operating conditions. Thus, the system hardware is also bound to fail at some point of time. This failure however, must be avoided at all costs, especially during the execution of any mission. It is therefore necessary to evaluate reliability of vehicles before launching it in an operation. A vehicle fleet manager is interested to know the reliability of his fleet of vehicles along with the logistic specific requirements for success mission. Estimation of reliability helps in not only evaluating the reliability of the fleet but also provides valuable inputs about weak link areas which can be improved upon by employing various maintenance techniques.

*Corresponding author's email: hari_delhi82@yahoo.com

Terminology and Notation

ANN	Artificial Neural Network
MAPE	Mean Absolute Percentage Error
RAM	Reliability, Availability and Maintainability
RDBMS	Relational Data Base Management System
SLP	Single Layer Perceptron
$Q(t)$	cumulative density function of the distribution
$R(t)$	Reliability
N	total fleet of vehicle population
c	intercept on y axis
i	rank order
k	number of class intervals
n_e	expected number of failures
n_i	number of failures
n_o	Observed number of failures
p_i	Probability of failure in i^{th} class
t	mid-class interval time in km
w	number of input (or output) neurons
β	shape parameter
γ	scale parameter
η	location parameter
χ^2	chi-square

2. Justification for using ANN

ANNs are computational modelling tool that has found extensive acceptance in many disciplines for modelling complex real-world problems. ANNs may be defined as structures comprised of densely inter-connected, adaptive, simple processing elements (called artificial neurons or nodes) that are capable of performing massively parallel computations for data processing and knowledge representation (Kumar *et. al.* [1], Park *et. al.* [2]). The attractiveness of ANNs comes from the remarkable information processing characteristics of the biological system such as non-linearity, high parallelism, robustness, fault and failure tolerance, learning, ability to handle imprecise and fuzzy information and their capability to generalise. One can find the details of structure of ANN in Basheer *et. al.* [5]. The best fitting statistical distribution may not be able to give a fit, such that the anticipated casualties calculated in a mission are within desired limits. It may give very optimistic as well as pessimistic predictions depending upon the extent of fit at various points of time. In such cases, the use of ANN is recommended to obtain the observed reliability values as output from the inputs. The inputs are normally taken as sequential numbers from *1* to *n*. However, that would amount to assuming that the best fitting distribution has a linear character, which would not be correct in the present context. Hence, the inputs can be taken as the reliability values from the best fitting probability distribution. The ANN can be used as a tool for reduction of difference between observed and perceived values (called errors). However, what must be kept in mind is that ANN has the uncanny ability to be trained on and almost completely replicate training data used for their learning. This implies that if the training data itself has a noise component, the ANN will learn that too, which is undesirable. This is called over-fitting of the curve on which the ANN is being trained. To overcome this problem, efforts should be taken to ensure that the data presented to the distributions and that presented to the ANN should be free from errors as far as possible.

3. Training the ANN

The ANN uses steepest descent rule for training the ANN. ANN training is an iterative process, where in the metric representing the performance of the ANN is measured every iteration and is steered in a direction to minimise or maximise the measure of performance as the case may be. For the first iteration, arbitrary weights are assigned to each of the connections. In the present case, constant weights were used for the first iteration. These weights were calculated using the formula:

Initial Weight = (Sum of Expected Outputs / Sum of Inputs)/ Number of Observations

Generally, the error on training data decreases indefinitely with increasing number of hidden nodes or training cycles, as shown in Fig. 1. The initial large drop in error is due to learning, but the subsequent slow reduction in error may be attributed to:

- Network memorisation resulting from the excessively large number of training cycles used.
- Over-fitting due to the use of a large number of hidden nodes.

During ANN training, the error on test subsets is monitored which generally shows an initial reduction and a subsequent increase due to memorisation and overtraining of the trained ANN. The final (optimal) neural network architecture is obtained at the onset of the increase in test data error. Other error metrics may be used and may perform equally well in terms of optimising network structure (Twomey and Smith [7]).This is based on the premise that there is only one input and output also the first weight would be the final weight. Subsequently, the weights are adjusted in subsequent iterations.

The convergence/stopping criteria used for assessing the learning or training of the ANN is the value of MAPE between the target values and the ANN output values at each output neuron. Another convergence/stopping criterion used for training ANN is the sum-of-squared-errors. MAPE is the mean of the absolute difference between the observed and forecast values, expressed as a percentage of the observed values (Cavalieri et. al. [6]):

$$\text{MAPE}=(1/n)\sum_{i=1}^{n}[(Observed\ Value_i - Forecast\ Value_i)/Observed\ Value_i]*100$$

In present case MAPE used as convergence criteria for ANN training. Normally, the MAPE from the first to the second iteration increases, hence the learning rate needs to be decreased and otherwise the errors will cross over the x-axis in subsequent iteration. Basically, an increasing error between iterations is an indicator of the unsuitability of the current learning rate vis-à-vis the current set of weights between nodes. Controlling of ANN learning rate in this manner is referred to as association of a momentum factor with the learning rate. This momentum decelerates the learning in the uphill direction (when the errors increase between iterations) and accelerates learning in the downhill direction (when errors decrease between iterations). The momentum factor, therefore, stabilises the weights when the system tends to oscillate.

4. Reliability Predictions using Trained ANN

Once the ANN is trained, it is be used for finding out the expected values of the reliability using the predictions obtained by applying the best fitting model at future points of time. There are two methods of presenting data to the ANN, the batch method and the in-line method. If the batch method is used, then the number of forecasts presented to the ANN is equal to the number of time periods for which expected values are required and removing an equal number of historical inputs. The limitation of this method in the present application is that the maximum number of expected values that can be found is less than or equal to the number of neurons in the input layer due to the number of weights calculated. In the in-line mode of data presentation, on the other hand, one new data point at a time is presented to the ANN after removing the oldest data point. This can be continued indefinitely. If single prediction is to be made at one time, then in-line method is preferred over batch method. However, for prediction of a series of values in time series, the batch method is used due to lesser amount of effort involved. A SLP model has been used for refinement of results predicted for Weibull distribution by inserting ANN layer. The inputs are the reliability values obtained using Weibull distribution and the outputs are the observed reliability values obtained from the grouped data at each of first ten mid-point values of *t*. The number of input and output neurons or the nodes is ten. A RDBMS which provides a ready-made framework for storing and handling data in the form of table structures is used for developing ANN software. The database structure needed is designed for the purpose.

5. Methodology

- Collection, sorting and grouping of failure data.
- Estimation of distribution parameters and carrying out goodness of fit test.
- Justification for using ANN and its training using MAPE
- Reliability prediction using Weibull distribution and trained ANN.
- Comparing observed reliability with predicted reliability using Weibull distribution and trained ANN.

6. Case Study

A case study to illustrate the methodology described above is discussed below. The grouped data indicating the failure frequency in various kilometre-age intervals of vehicles is given in Table 1. The sorting and grouping of the raw data has been carried out using Sturge's Heuristic $k = (1+3.3\log_{10}N)$, 'N' in this case is 3046.

6.1 Identification of the Statistical Distribution

The probability plot for Weibull distribution has been used for finding the distribution parameters. The expected frequencies of failure are then calculated using the distribution parameters to estimate transport vehicle fleet mission reliability (Prasad *et.al.* [8]). Table 1 shows the calculation of the *ln t* and *lnln1/(1-Q(t))* from the grouped data. The value of '*i*' has been calculated using Bernard's empirical formula. The *ln t* and *lnln1/(1-Q(t))*, assuming η=0, have been calculated. The Fig. 2 is plotted using the values of *ln t* and *lnln1/(1-Q(t))* from Table 1 which is used for evaluating the values of β and γ for the Weibull distribution (Misra [3]).

From the Fig. 2 β = 1.97, c = -15.8 and γ is:

$$\gamma = e^{-c/\beta} = e^{-(-15.8)/1.97} = 3042.104 \text{ kms}$$

From Fig. 1 it can be seen that probability plot has given a very good-fit. However the goodness-of-fit can only be ascertained using chi-square goodness-of-fit test.

Table 1: Probability Plot Variables

S. No	t	n_i	$\ln t$	i	$Q(t)$	$\ln\ln i(1\text{-}Q(t)$
1	200	22	5.298	11	0.004	-5.65
2	600	128	6.397	86	0.028	-3.56
3	1000	205	6.908	252.5	0.083	-2.45
4	1400	275	7.244	492.5	0.162	-1.74
5	1800	319	7.496	789.5	0.259	-1.20
6	2200	339	7.696	1119	0.367	-0.78
7	2600	334	7.863	1455	0.478	-0.43
8	3000	305	8.006	1775	0.582	-0.14
9	3400	275	8.132	2065	0.678	0.124
10	3800	232	8.243	2318	0.761	0.358
11	4200	184	8.343	2526	0.829	0.569
12	4600	120	8.434	2678	0.879	0.748
13	5000	124	8.517	2800	0.919	0.922
14	5400	79	8.594	2902	0.952	1.113
15	5800	41	8.666	2962	0.972	1.274
16	6200	43	8.732	3004	0.986	1.448
17	6600	21	8.795	3036	0.996	1.724

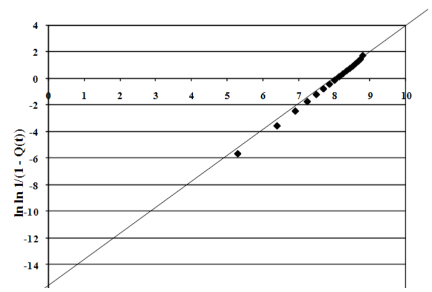


Figure 1: Probability Plot

The number of parameters estimated in the present study is two. Therefore, the number of degrees of freedom for the test is 14. The critical value of χ^2 at 5% significance level and 14 degrees of freedom is 23.69, which is more than the calculated value. Hence there is a close agreement between the observed and expected frequencies which suggests that the Weibull distribution provides a "good fit" (Ebeling [4]).

Table 2 : Chi-Square Test for Weibull Distribution

S. No	t	n_o	$Q(t)$	p_i	n_e	χ^2
1	200	22	0.0046791	0.0046791	14.25240	4.21159
2	600	128	0.0400190	0.0353400	107.64558	3.84876
3	1000	205	0.1057091	0.0656901	200.09198	0.12039
4	1400	275	0.1948930	0.0891839	271.65414	0.04121
5	1800	319	0.2992886	0.1043956	317.98898	0.00321
6	2200	339	0.4102783	0.1109897	338.07453	0.00253
7	2600	334	0.5199731	0.1096948	334.13050	0.00005
8	3000	305	0.6220214	0.1020482	310.83895	0.10968
9	3400	275	0.7120519	0.0900306	274.23306	0.00214
10	3800	232	0.7877438	0.0756919	230.55757	0.00902
11	4200	184	0.8485904	0.0608466	185.33868	0.00967
12	4600	120	0.8954711	0.0468807	142.79873	3.63996
13	5000	124	0.9301530	0.0346819	105.64105	3.19053
14	5400	79	0.9548224	0.0246694	75.14299	0.19798
15	5800	41	0.9717125	0.0168901	51.44711	2.12144
16	6200	43	0.9828528	0.0111403	33.93336	2.42251
17	6600	21	0.9899365	0.0070838	21.57711	0.01544

$\Sigma\chi^2 = 19.946$

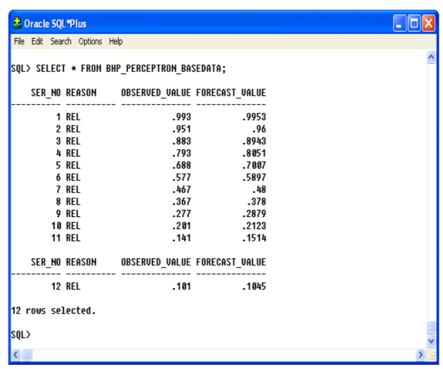
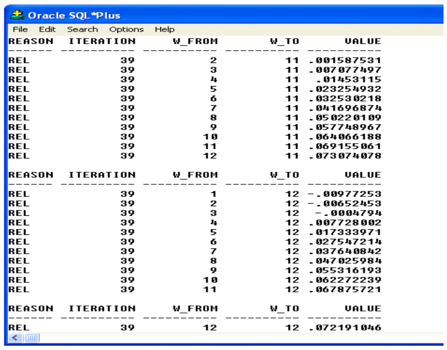

```
SQL> SELECT * FROM BHP_PERCEPTRON_BASEDATA;

    SER_NO REASON     OBSERVED_VALUE FORECAST_VALUE
---------- ---------- -------------- --------------
         1 REL                  .993          .9953
         2 REL                  .951            .96
         3 REL                  .883          .8943
         4 REL                  .793          .8051
         5 REL                  .688          .7007
         6 REL                  .577          .5897
         7 REL                  .467            .48
         8 REL                  .367           .378
         9 REL                  .277          .2879
        10 REL                  .201          .2123
        11 REL                  .141          .1514

    SER_NO REASON     OBSERVED_VALUE FORECAST_VALUE
---------- ---------- -------------- --------------
        12 REL                  .101          .1045

12 rows selected.

SQL>
```


Figure 2 : Data Presented for Perceptron Training

6.2 Reliability Prediction using Trained ANN

The data presented to the perceptron for training is shown in Fig. 2. The inputs *i.e.*, the Weibull reliability values have been calculated using the expression, $R(t) = e^{-(t/\gamma)^{\beta}}$, whereas the values of γ and β have been calculated, as 3042.104 kms and 1.97 respectively.

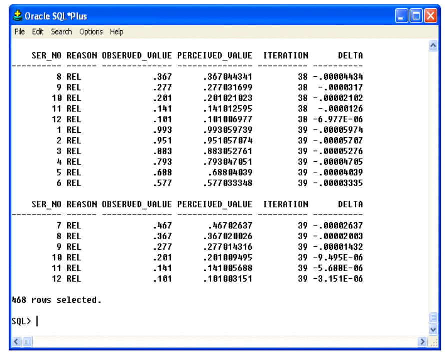


REASON	ITERATION	W_FROM	W_TO	VALUE
REL	39	2	11	.001587531
REL	39	3	11	.007077497
REL	39	4	11	.01453115
REL	39	5	11	.023254932
REL	39	6	11	.032530218
REL	39	7	11	.041696874
REL	39	8	11	.050220109
REL	39	9	11	.057748967
REL	39	10	11	.064066188
REL	39	11	11	.069155061
REL	39	12	11	.073074078

REASON	ITERATION	W_FROM	W_TO	VALUE
REL	39	1	12	-.00977253
REL	39	2	12	-.00652453
REL	39	3	12	-.0004794
REL	39	4	12	.007728002
REL	39	5	12	.017333971
REL	39	6	12	.027547214
REL	39	7	12	.037640842
REL	39	8	12	.047025984
REL	39	9	12	.055316193
REL	39	10	12	.062272239
REL	39	11	12	.067875721

REASON	ITERATION	W_FROM	W_TO	VALUE
REL	39	12	12	.072191046

Figure 3: Weights Generated by ANN Program



SER_NO	REASON	OBSERVED_VALUE	PERCEIVED_VALUE	ITERATION	DELTA
8	REL	.367	.367044341	38	-.00004434
9	REL	.277	.277031699	38	-.0000317
10	REL	.201	.201021023	38	-.00002102
11	REL	.141	.141012595	38	-.0000126
12	REL	.101	.101006977	38	-6.977E-06
1	REL	.993	.993059739	39	-.00005974
2	REL	.951	.951057074	39	-.00005707
3	REL	.883	.883052761	39	-.00005276
4	REL	.793	.793047051	39	-.00004705
5	REL	.688	.68804039	39	-.00004039
6	REL	.577	.577033348	39	-.00003335

SER_NO	REASON	OBSERVED_VALUE	PERCEIVED_VALUE	ITERATION	DELTA
7	REL	.467	.46702637	39	-.00002637
8	REL	.367	.367020026	39	-.00002003
9	REL	.277	.277014316	39	-.00001432
10	REL	.201	.201009495	39	-9.495E-06
11	REL	.141	.141005688	39	-5.688E-06
12	REL	.101	.101003151	39	-3.151E-06

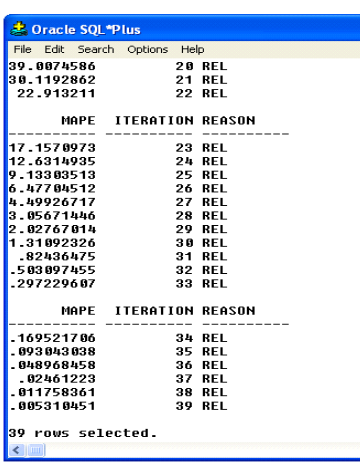
468 rows selected.

SQL> |

Figure 4: Error between Observed and Perceived Value

The outputs, the observed reliabilities at various points of time have been calculated from grouped data by dividing the number of survivors from the total number in service *i.e.,* 3046. Weights generated by perceptron program in last iteration are shown in Fig. 3. The program generated 5616 rows in 39 iterations till the value of observed value and perceived/forecasted value by the perceptron program are near equal. The error between the observed and the perceived values is calculated by the perceptron program. The SLP took thirty nine iterations with a learning rate of two to reduce the error between observed value and perceived value/forecast value by the perceptron program to bring the error of MAPE less than 0.01%. The first two and the last two iterations are shown in Figure. 4.

The differences between observed and perceived values (error) are then used for calculating the MAPE. The MAPE values plotted against the number of iteration gives the learning curve. The MAPE values generated by perceptron program are shown in Fig. 5 and the learning curve is given in Fig. 6. As can be seen from Fig. 6, the MAPE from the first to the second iteration increases, hence the learning rate needs to be decreased, otherwise the errors will cross-over the *x*-axis in subsequent iterations, with increasing amplitude. Basically, an increasing error between iterations is an indicator of the unsuitability of the current learning rate vis-à-vis the current set of weights between nodes.



39.0074586	20	REL
30.1192862	21	REL
22.913211	22	REL

MAPE	ITERATION	REASON
17.1570973	23	REL
12.6314935	24	REL
9.13303513	25	REL
6.47704512	26	REL
4.49926717	27	REL
3.05671446	28	REL
2.02767014	29	REL
1.31092326	30	REL
.82436475	31	REL
.503097455	32	REL
.297229607	33	REL

MAPE	ITERATION	REASON
.169521706	34	REL
.093043038	35	REL
.048968458	36	REL
.02461223	37	REL
.011758361	38	REL
.005310451	39	REL

39 rows selected.

Figure 5: MAPE Values

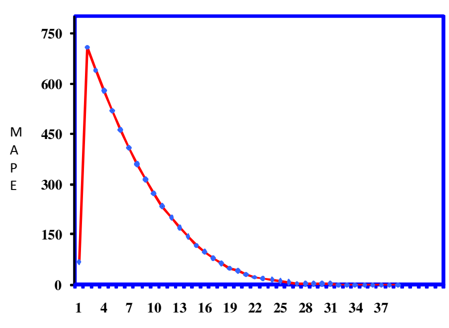


Figure 6: Learning Curve

Estimated reliabilities have been calculated using $R(t) = e^{-(t/\gamma)^{\beta}}$, with parameters γ and β determined above. The observed reliabilities have been calculated from grouped data by dividing the number of survivors from the total number of vehicles used in mission, of mixed vintage which were selected using multiply censored data using purposively (judgemental) sampling. The predicted reliability values, when next seven Weibull reliability values are fed to the SLP in batch mode are given in Table 3 and Fig. 7. Reliability values shown in bold are the predicted values for which no failure data exists, and are refined estimates compared to pure Weibull reliability estimates. A plot of Weibull, observed and predicted reliabilities, is given in Fig. 7. The plot shows that the ANN and observed reliabilities are closer as compared to Weibull and observed ones.

Table 3 : Results of Reliability Predictions using SLP

S. No	t	Weibull Reliability	Observed Reliability	Predicted Reliability using ANN
1	200	0.995320	0.9930	0.993062
2	600	0.959980	0.9510	0.951060
3	1000	0.894290	0.8830	0.883056
4	1400	0.805106	0.7930	0.793050
5	1800	0.700711	0.6880	0.688043
6	2200	0.589721	0.5770	0.577037
7	2600	0.480026	0.4670	0.467030
8	3000	0.377978	0.3670	0.367024
9	3400	0.287948	0.2770	0.277020
10	3800	0.212256	0.2010	0.201019
11	4200	0.151409	0.1410	0.141019
12	4600	0.120452	0.1201	0.120120
13	5000	0.069846	0.1170*	0.117030
14	5400	0.045177	0.0910*	0.091280
15	5800	0.028287		0.067380
16	6200	0.017146		0.046081
17	6600	0.010063		0.028101
18	7000	0.005718		0.013911
19	7400	0.003146		0.004460

* Data not used for ANN but used for validation

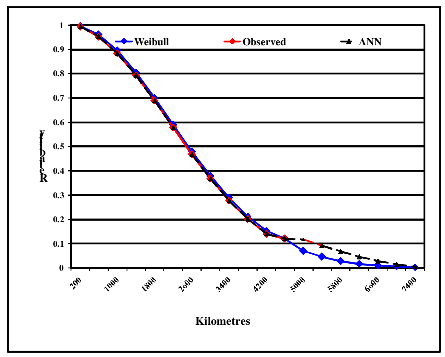


Figure 7 : Weights Generated by ANN Program

6.3 Analysis of Results

After analysis of the results, the important findings of this paper are as under:

- This paper provides a methodology using ANN (SLP model) and for prediction of vehicle reliability.
- By using RDBMS (Oracle), training of ANN has been successfully carried out to predict the reliability of vehicle. It is found that by adopting this method the difference between the observed and predicted reliability of vehicle is very less, which is very difficult to achieve using probability distributions.
- It is observed that the difference between observed and predicted values of vehicle reliability for initial and subsequent values of t km is gradually increasing for Weibull probability distribution. However for trained ANN predicted value is almost same as the observed values. For t=200 km, the error between observed and predicted reliability using Weibull probability distribution is 0.233% while for t=5400 km the error is 50.35%. However for t=200 km and t=5400 km with trained ANN the error is 0.00624371% and 0.30769231% respectively, as can be seen in Table 3 and Fig. 7.

- As prediction of reliability using trained ANN is accurate, this technique may prove to be very useful and effective in predicting the reliability of critical equipments/systems.

7. Conclusions

The paper proposes a procedure for refinement of reliability predictions of vehicles, estimated by application of statistical principles on field failure data, is described. Use of ANN for this purpose is emphasized. The addition of the ANN layer over the wide range of statistical distributions, for improving the accuracy of the predicted values of reliability is demonstrated using a single layer perceptron model. Back propagating ANN models can also be used in the same conceptual framework of proposed model. However the computational effort would increase manifold with addition of hidden layers. The implementation of the SLP has been achieved using RDBMS. The suggested procedure thus provides a useful decision support tool, in the hands of vehicle fleet managers to work out the vehicle reliability and the expected number of failures during a mission.

Acknowledgement: Authors would like to thank the anonymous referees who helped improve the paper.

References

[1] Kumar, K. *Effect of ranking on reliability prediction using Weibull Analysis.* Reliability Engineering and System Safety, 1989; 24(3): 257-65.

[2] Park, K.S., and Y.K. Yoo. *Reliability Apportionment for Phased Mission Oriented Systems.* Reliability Engineering and System Safety, 1990; 27(3): 357-364.

[3] Misra, K.B. *Reliability Analysis and Prediction.* Elsevier Science Publishers, Netherlands, 1992.

[4] Ebeling, C.E. *Introduction to Reliability and Maintainability Engineering.* Tata Mc Graw Hill, New Delhi, 2000.

[5] Basheer, I.A., and M. Hajmeer. *Artificial neural networks: fundamentals, computing, design, and application.* Journal of Microbiological Methods, 2000; 43: 3-31.

[6] Cavalieri, S., P. Maccarrone, and R. Pinto. *Parametric versus neural network models for the estimation of production costs: A case study in the automotive industry.* International Journal of Production Economics, 2003; 165-177.

[7] Twomey, J. and A. Smith. *Artificial Neural Networks for Civil Engineers: Fundamentals and Applications.* ASCE, 1997; 44-64.

[8] Prasad, B.H., S.B. Singh and M.S. Bhat. *Vehicular Fleet Reliability Estimation: A Case Study.* International Journal of Performability Engineering, 2009; 5(3): 235-242.

B. Hari Prasad is serving in Indian Armed Forces since last 25 years. He received his B. Tech. from Punjab Engineering College Chandigarh, M. Tech. from Indian Institute of Technology (IIT) Delhi and Ph.D. in Reliability Engineering from Jawaharlal Nehru Technological University (JNTU), Hyderabad, India. His M. Tech project was under Dr Prem Vrat (then Director, IIT Delhi), which was adjudged as the best. He has published several research papers in national and international journals.

P. Bhattacharjee currently working as Director Grade-I with Defence research and Development Laboratory at Hyderabad. He has published several papers in national and international journals.

A. Venu Gopal has done his Ph.D. in manufacturing from IIT Delhi. He is currently working as Associate Professor in Mechanical Engineering Department, National Institute of Technology, Warangal, India for the past 22 years. He has published several research papers in national and international journals.