

Progressive prediction method for failure data with small sample size

WANG Zhi-hua¹, FU Hui-min¹, LIU Cheng-rui²

(1. Research Center of Small Sample Technology,
Beijing University of Aeronautics and Astronautics, Beijing 100191, China;

2. Science and Technology on Space Intelligent Control Laboratory,
Beijing Institute of Control Engineering, Beijing 100191, China)

Abstract: The small sample prediction problem which commonly exists in reliability analysis was discussed with the progressive prediction method in this paper. The modeling and estimation procedure, as well as the forecast and confidence limits formula of the progressive auto regressive (PAR) method were discussed in great detail. PAR model not only inherits the simple linear features of auto regressive (AR) model, but also has applicability for nonlinear systems. An application was illustrated for predicting the future fatigue failure for Tantalum electrolytic capacitors. Forecasting results of PAR model were compared with auto regressive moving average (ARMA) model, and it can be seen that the PAR method can be considered good and shows a promise for future applications.

Key words: failure data forecast; system reliability; small sample; progressive prediction;
nonlinear system

CLC number: TB114; O211

Document code: A

Introduction

In system reliability analysis, few would argue about the need and importance of forecasting as decisions makers are interested in estimating the future occurrences of system failure for resource planning, inventory management, developing realistic policies for age replacement and logistic support^[1]. However, small sample prediction problem are often encountered in engineering, since it can not be allowed to increase sample number for getting large amounts of field data in many reliability experiment occasions^[2]. The difficulty of failure data prediction with small sample size lies in that analyzer can not determine whether it has value to analyze and forecast or not, but can only speculate the approxi-

mate trend of the sequence based on the scanty and poor information the single scatter plot only provided^[3].

Many methods have been developed to solve that problem over the years. Neural network can almost predict all fault sequences, and the non-linear mapping ability of neural networks enables it to be applied to nonlinear systems^[4-5]. However, in practical engineering, neural networks structure, including hidden nodes, hidden layers and the role of functions are all determined by experience. In addition, the extra related error can not be analyzed. Regression analysis method based on support vector machine can fit the development trend for the sequence and obtain linear and nonlinear interval, however, it can not provide correlative information from

Received: 2011-03-04; **revision received:** 2011-05-11

Foundation items: Supported by Fanzhou Science and Research Foundation for Young Scholars (Grant No. 20100511)

E-mail: wangzhihua@buaa.edu.cn

sample data and can not consider the noise influence^[6]. Grey prediction method can establish grey model and fit exponential growth sequence trend well by means of sorting failure data order and accumulating them^[7-8]. Although this seemed to go one step forward in non-linear prediction, it is also powerless to fit sequence correctly and make effective prediction, because it is only a curve fitting method that can not fetch certain identification information of sequence from angle of correlation.

Auto regressive moving average (ARMA) model can express the future value of stationary sequence as a linear combination of the past values and random noise based on minimum square error^[9]. In this method, the present behavior of sequence could be inferred from its past behavior, and its successful application in fault data prediction fully demonstrates this advantage^[10]. However, it also has the limitation that at least 50 and preferably 100 observations or more should be used, and this method can not be applied to nonlinear systems^[11]. Fu proposed a new method called progressive auto regressive (PAR) model in Ref. [12]. This method could make full use of the linear advantage of ARMA model, and expand its applicability to non-linear prediction problem with small sample size.

The establishment of PAR model was analyzed with small sample failure data in this paper. In section1, we start from the basic idea of progressive failure data prediction with small sample size, and then we present the prediction model and explain the estimation of model parameters. Section2 contains application to a group of small sample failure data according to modeling principle and method. Multi-step ahead predictions and 0.95 confidence intervals were given. In Section3, we close this paper with a discussion and conclusion.

1 Progressive prediction method for failure data

If x_t , $t = 1, 2, \dots, n$, is a failure series,

where n is a considerably sample size, the establishment of PAR model and estimation of model parameters will be described in great detail. The 1st-step-ahead prediction model can be established based on initial observational failure sample $x_1, x_2, \dots, x_n$ as

$$\begin{cases} \Phi_1(L)x_t = \alpha_{10} + \epsilon_{1t} \\ \epsilon_{1t} \sim \text{NID}[0, \sigma_{1\epsilon}^2] \end{cases} \quad (1)$$

where $\Phi_1(L) = 1 - \varphi_{11}L - \dots - \varphi_{1p_1}L^{p_1}$ specifies the 1st-step-ahead PAR lag-polynomial with model order p_1 , and α_{10} is a constant of the 1st-step-ahead PAR model. Then we can obtain the 1st-step-ahead prediction result by

$$\hat{x}_{n+1|n} = \alpha_{10} + \varphi_{11}x_n + \dots + \varphi_{1p_1}x_{n+1-p_1}$$

Then we consider the failure data set of $x_1, x_2, \dots, x_n, \hat{x}_{n+1|n}$, and roll the sample window on $x_2, x_3, \dots, x_n, \hat{x}_{n+1|n}$, and the 2nd-step-ahead PAR model can be built as

$$\begin{cases} \Phi_2(L)x_t = \alpha_{20} + \epsilon_{2t} \\ \epsilon_{2t} \sim \text{NID}[0, \sigma_{2\epsilon}^2] \end{cases} \quad (2)$$

where $\Phi_2(L) = 1 - \varphi_{21}L - \dots - \varphi_{2p_2}L^{p_2}$ specifies the 2nd-step-ahead PAR lag-polynomial with model order p_2 , and α_{20} is a constant of the 2nd-step-ahead PAR model. And we can obtain the 2nd-step-ahead prediction result by

$$\hat{x}_{n+2|n} = \alpha_{20} + \varphi_{21}\hat{x}_{n+1|n} + \varphi_{22}x_n + \dots + \varphi_{2p_2}x_{n+2-p_2}$$

Analogically, for l th-step-ahead settlement prediction, we roll the data window on sample set $x_t^*, x_{t+1}^*, \dots, x_{n+t-1}^*$, where

$$x_k^* = \begin{cases} x_k & k \leq n \\ \hat{x}_{k|n} & k > n \end{cases} \quad (3)$$

then the l th-step-ahead PAR equation can be established as

$$\begin{cases} \Phi_l(L)x_t = \alpha_{l0} + \epsilon_{lt} \\ \epsilon_{lt} \sim \text{NID}[0, \sigma_{l\epsilon}^2] \end{cases} \quad (4)$$

where $\Phi_l(L) = 1 - \varphi_{l1}L - \dots - \varphi_{lp_l}L^{p_l}$ specifies the l th-step-ahead PAR lag-polynomial with model order p_l , and α_{l0} is a constant of the l th-step-ahead PAR model.

The determination of model order p_l for the l th-step-ahead PAR equation (4) can be a-

chieved by Akaike information criterion (AIC) rule^[13].

According to least-squares method, the constant of the l th-step-ahead PAR model α_{l0} can be calculated by

$$\alpha_{l0} = \bar{x}_{l0} - \sum_{i=1}^{p_l} \varphi_{li} \bar{x}_{li} \quad (5)$$

where mean $\bar{x}_{li}$ with $i = 1, 2, \dots, p_l$ can be obtained by the following formula

$$\bar{x}_{li} = \frac{1}{n - p_l} \sum_{t=p_l+1}^{n+l-1} x_{t-i}^* \quad i = 1, 2, \dots, p_l \quad (6)$$

where x_t^* is given by equation (3). The autoregressive coefficients φ_{li} with $i = 1, 2, \dots, p_l$ can be determined by

$$\begin{pmatrix} \varphi_{l1}, \varphi_{l2}, \dots, \varphi_{lp_l} \end{pmatrix}^T = \left(\mathbf{L}_{l\varphi\varphi} \right)_{p_l \times p_l}^{-1} \left(\mathbf{L}_{l\varphi} \right)_{p_l \times 1} \quad (7)$$

where $\left(\mathbf{L}_{l\varphi\varphi} \right)_{p_l \times p_l} = \left(S_{lij} \right)_{p_l \times p_l}$ is a p_l -order matrix, and $\left(\mathbf{L}_{l\varphi} \right)_{p_l \times 1} = \left(S_{li1}, S_{li2}, \dots, S_{lip_l} \right)^T$ is a p_l -order column vector. The elements can be determined through the following calculation

$$S_{lij} = \sum_{t=p_l+1}^{n+l-1} (x_{t-i}^* - \bar{x}_{li})(x_{t-j}^* - \bar{x}_{lj}) \quad i, j = 1, 2, \dots, p_l \quad (8)$$

$$S_{li} = \sum_{t=p_l+1}^{n+l-1} (x_t^* - \bar{x}_{l0})(x_{t-i}^* - \bar{x}_{li}) \quad i = 1, 2, \dots, p_l \quad (9)$$

And the estimator of $\sigma_{l\epsilon}^2 = \text{Var}(\epsilon_{lt})$ can be obtained by

$$\sigma_{l\epsilon}^2 = \frac{1}{n + l - p_l - 2} \sum_{t=p_l+1}^{n+l-1} \left(x_t^* - \alpha_{l0} - \sum_{i=1}^{p_l} \varphi_{li} \bar{x}_{li} \right)^2 \quad (10)$$

Then the l th-step-ahead failure prediction value $\hat{x}_{n+l|n}$ can be determined by

$$\hat{x}_{n+l|n} = \alpha_{l0} + \sum_{i=1}^{p_l} \varphi_{li} x_{n+l-i|n}^* \quad (11)$$

And its mean square error (MSE) $\sigma_{n+l|n}^2$ can be obtained by

$$\sigma_{n+l|n}^2 = E(x_{n+l} - \hat{x}_{n+l|n})^2 = \sum_{i=1}^l \lambda_{li}^2 \sigma_{(n+l-i)}^2 \quad (12)$$

The two-side $100\gamma\%$ confidence interval can be calculated by

$$\hat{x}_{n+l|n} - u_{(1+\gamma)/2} \sigma_{n+l|n} \leq x_{n+l} \leq \hat{x}_{n+l|n} + u_{(1+\gamma)/2} \sigma_{n+l|n} \quad (13)$$

where $\varphi_{li} = 0$, when $i > p_l$, and

$$\begin{cases} \lambda_{kl} = 1 & k = 1, 2, \dots, l \\ \lambda_{ki} = \sum_{j=1}^{i-1} \varphi_{kj} \lambda_{(k-j)(i-j)} & k = 2, 3, \dots, l; \\ & i = 2, 3, \dots, k \end{cases} \quad (14)$$

The failure prediction result obtained by the proposed PAR model shows best unbiased property, which can be proved by mathematical induction method from the least square error (LSE) forecasting method in Ref. [14].

2 Application

The application we illustrate in this section is the use of PAR models to forecast the future fatigue failure for Tantalum electrolytic capacitors. Fig. 1 shows the scatter plot, and the fatigue data were under the stress of 65 V.

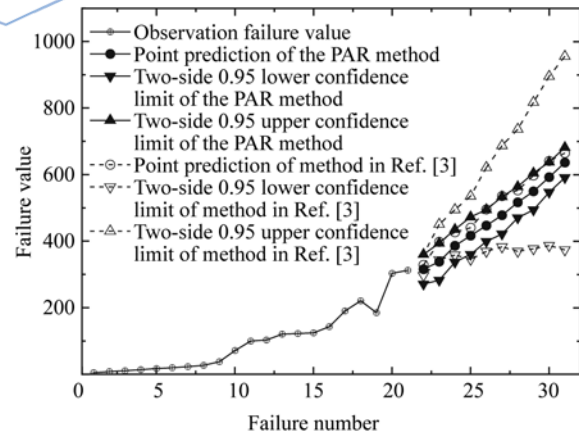


Fig. 1 Failure observations and prediction results

Let x_t , $t = 1, 2, \dots, n$, indicate the observational failure data series obtained, where $n = 21$ corresponds to the sample size. The PAR model construction results for the future 10 steps ahead at $l = 1, 2, \dots, 10$, are summarized in Table 1. Prediction by the ARMA model with parameter

Table 1 Parameter results of PAR model

Step l	Model parameter				
	p_l	α_0	φ_{l1}	φ_{l2}	φ_{l3}
1	2	11.01	0.309	0.226	0.751
2	4	11.57	1.035		
3	1	17.75	0.483	0.200	0.455
4	1	12.12	1.046		
5	1	22.75	0.456	0.194	0.471
6	2	15.07	1.035		
7	5	30.55	0.416	0.190	0.487
8	3	18.49	1.027		
9	2	25.15	0.399	0.222	0.488
10	2	24.54	0.393	0.222	0.496

derivation method of particle swarm optimization (PSO) algorithm based on maximum likelihood estimation (MLE) was executed by Ref. [3]. The point forecasts and two-side 0.95 confidence limits of the PAR model and method in Ref. [3] are summarized in Table 2 and also visually illustrated in Fig. 1.

As we know in interval prediction that, the size of confidence interval reflects the prediction accuracy under given confidence level. Smaller confidence interval means higher prediction accuracy under given confidence level. Consequently, we can conclude from the prediction results in Fig. 1 and Table 2 that, accuracy of the PAR method is better than ARMA model in Ref. [3].

Table 2 10 steps ahead prediction results out of series

Step l	Point prediction		Two-side 0.95 upper confidence limit		Two-side 0.95 lower confidence limit		Length of confidence interval	
	PAR method	Method in Ref. [3]	PAR method	Method in Ref. [3]	PAR method	Method in Ref. [3]	PAR method	Method in Ref. [3]
1	316.1	329.2	360.6	360.9	271.7	297.4	88.9	63.5
2	338.8	398.2	394.7	451.5	283.0	344.8	111.6	106.7
3	386.9	427.1	436.0	495.4	337.8	358.7	98.3	136.7
4	417.0	441.1	473.3	536.1	360.6	346.1	112.7	190
5	447.5	496.0	495.5	621.9	399.5	370	96.0	251.9
6	478.3	535.8	534.2	688	422.4	383.5	111.9	304.5
7	517.4	553.7	564.1	737.2	470.8	370.1	93.3	367.1
8	549.9	597.7	605.2	817.5	494.5	377.7	110.7	439.8
9	592.9	641.5	638.2	895.7	547.5	387.3	90.7	508.4
10	636.8	665.2	682.1	955.8	591.5	374.7	90.6	581.1
Average length of confidence interval							100.5	295.0

3 Conclusions

In reliability field, modeling and prediction study of small sample failure data has shown clear and significant sense, because it plays an important role in reducing reliability test time and cost.

The principle and method to establish PAR

model for failure data with small sample size were discussed in this paper. Application showed the feasibility and effectiveness of PAR model for the failure prediction problem with small sample size.

References:

- [1] Ho S L, Xie M. The use of ARIMA models for reliability

- forecasting and analysis[J]. Computers and Industrial Engineering, 1998, 35(1-2): 213-216.
- [2] ZOU Xinyao, YAO Ruihe. Small sample statistical theory and IC reliability assessment[J]. Control and Decision, 2008, 23(3): 241-244. (in Chinese)
- [3] HUANG Jianguo, LUO Hang, LONG Bing, et al. Prediction research about small sample failure data based on AR-MA model[C]// Proceedings of IEEE Circuits and Systems International Conference on Testing and Diagnosis. China, Chengdu: IEEE, 2009: 1-6.
- [4] Connor J T, Martin D, Atlas L E. Recurrent neural networks and robust time series prediction[J]. IEEE Transactions on Neural Networks, 1994, 5(2): 240-254.
- [5] Rietman E A, Beachy M. A study on failure prediction in a plasma reactor[J]. IEEE Transactions on Semiconductor Manufacturing, 1998, 11(4): 670-680.
- [6] Hong D H, Hwang C. Interval regression analysis using quadratic loss support vector machine[J]. IEEE Transactions on Fuzzy Systems, 2005, 13(2): 229-237.
- [7] CHENG Huitao, HUANG Wenhui, JIANG Xingwei. Grey model based fault forecasting technique and its application in propulsion system of space[J]. Journal of Propulsion Technology, 1998, 19(3): 74-77. (in Chinese)
- [8] YANG Jiantian, YUE Weiliang. Application of grey model to machinery faults prediction[J]. Journal of Mechanical Strength, 2001, 23(3): 277-279. (in Chinese)
- [9] Box G E P, Jenkins G M, Reinsel G C. Time series analysis: forecasting and control[M]. San Francisco: Holden-day, 1976.
- [10] ZHAO Jie, XU Limei, LIU Liu. Equipment fault forecasting based on ARMA model[C]// Proceedings of the 2007 IEEE International Conference on Mechatronics and Automation. China, Harbin: IEEE, 2007: 3414-3518. (in Chinese)
- [11] Tseng F M, Tzeng G H, Yu H C, Yuan Benjamin J C. Fuzzy ARIMA model for forecasting the foreign exchange market[J]. Fuzzy Sets and Systems, 2001, 118(1): 9-19.
- [12] FU Huimin. Progressive autoregressive prediction method[J]. Journal of Mechanical Strength, 2006, 28(1): 34-39. (in Chinese)
- [13] Akaike H. A new look at the statistical model identification[J]. IEEE Transactions on Automatic Control, 1974, 19(6): 716-723.
- [14] Hamilton J D. Time series analysis[M]. New Jersey: Princeton University Press, 1994.